2018

Topics in bootstrap methods for survey sampling and spatially balanced design

Zhonglei Wang
Iowa State University

Topics in bootstrap methods for survey sampling and spatially balanced design

by

Zhonglei Wang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Jae Kwang Kim, Co-major Professor
Zhengyuan Zhu, Co-major Professor
Mark S. Kaiser
Chong Wang
Emily J. Berg

Iowa State University

Ames, Iowa

2018

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

ACKNOWLEDGEMENTS

I would like to thank my major professors, Dr. Zhengyuan Zhu and Dr. Jae Kwang Kim, for their valuable help through my research and life during my Ph. D. study. Their great enthusiasm in research and patience in guidance encourage me to finish and continue my research. I would also like to thank my committee members for their constructive and detailed suggestions for this work. I would like to express my thanks to the professors, staffs, and students in the Center for Survey Statistics & Methodology for their generous help during my Ph. D. study. Besides, I would like to thank Dr. Xiaogu Zheng and Dr. Song Xi Chen for their help during my early stage in statistics. At last, I would like to thank my family and friends for their kind support.

ABSTRACT

This dissertation consists of three parts. In the first part, we propose new bootstrap methods for three commonly used sampling designs, including the Poisson sampling, simple random sampling, and probability-proportional-to-size sampling. We show that the proposed bootstrap methods are second-order accurate and easy to be implemented in practice. Two simulation studies are conducted to compare the proposed bootstrap methods with the Wald method, and the proposed bootstrap methods outperform the Wald method in terms of coverage rate. It is well-known that a spatially balanced sample, which spread over the study domain well, can improve the estimation efficiency under dependent settings. In the second part, we propose to use a block bootstrap method to estimate the variance and make inference based on a sample generated by a one-per-stratum sampling design. We show the validity of the block bootstrap method and compare it with another commonly used sampling design theoretically. Simulation study shows that the block bootstrap method can provide valid variance estimator and inference for the one-per-stratum sampling design. Although there are many researches about spatially balanced sampling design, there are few discussing the spatio-temporal balanced sampling design. In the third part, we propose a spatio-temporal balanced sampling design to generate annual samples, such that the sample for each year is spatially balanced, and the one combining from consecutive years is also spatially balanced. We also propose design-based variance estimator for the estimates of annual status and annual change. The proposed sampling design is used in the National Resources Inventory rangeland on-site survey, and it shows that the proposed design performs better than the current design and estimators.

CHAPTER 1. OVERVIEW

As the computational power grows, bootstrap (Efron, 1979) becomes more and more popular since they are easy to implement and can provide more accurate estimates than the Wald method, which is based on the asymptotic normality; see Hall (1992) for details. Specific in survey sampling, there are some researches about bootstrap methods (Gross, 1980; Rao and Wu, 1988; Sitter, 1992a,b; Shao and Sitter, 1996). However, there is limitation in the existed bootstrap methods. First, most of them are proposed under a specific sampling design, that is, the stratified random sampling. Second, most of them are proposed for the variance estimation. Although some researchers (Rao and Wu, 1988) demonstrated that their methods achieve second-order accuracy, it is based on some known population quantities, which are hard to obtain in practice.

In the first project, we propose new bootstrap methods for three commonly used sampling designs, including the Poisson sampling, simple random sampling, and probability-proportional-to-size sampling, and we focus on the inference for the population total. Under very mild conditions, we show that the proposed bootstrap methods are second-order accurate, and we conduct two simulation studies to show that the proposed bootstrap methods are more preferable compared with the Wald method in terms of the coverage rate.

For environmental studies, it is desirable to obtain a spatially balanced sample, which spreads over the study domain well, to improve the estimation efficiency (Bellhouse, 1977). Stevens and Olsen (2004) proposed the generalized random tessellation stratified design, and they showed that the proposed design can generate a sample with better spatial balance compared with independent random sampling fand spatially stratified sampling. Grafström et al. (2012) proposed two local pivotal methods, which can generate a sample with better spatial balance compared with the generalized random tessellation stratified design. Furthermore, the local pivotal methods are more computational efficient than the generalized random tessellation stratified design, and they can be easily generalized to higher dimensional spaces. See Munholland and Borkowski (1996), Breidt (1995), Dunn and Harrison (1993), and Lister and Scott (2009) about other spatially balanced sampling designs. However, it is hard to obtain an unbiased variance estimator for most of the spatially balanced sampling designs.

In the second part, we propose to use the grid-based block bootstrap (Lahiri and Zhu, 2006) to obtain the variance estimator and make inference under a one-per-stratum sampling design. We show that the grid-based block bootstrap is valid under the one-per-stratum sampling design in weak dependent settings, and a simulation study shows that the performance of the block bootstrap method is good under different scenarios.

Although there are many researches about the spatially balanced sampling designs, there are few to study the spatio-temporal balanced sampling design, which generates annual samples such that the sample for each year is spatially balanced, and the one combined from consecutive years is also spatially balanced. In the third part, we propose a spatio-temporal balanced sampling design based on the local pivotal methods. We propose to use a regression estimator to estimate the annual status and change by borrowing information across samples for different years, and we also derive the corresponding design-based variance estimators. We applied the proposed sampling design to the National Resources Inventory rangeland on-site survey and showed that the proposed design is more efficient in estimating the annual quantities compared with the current sampling design and estimation methods.

References

Bellhouse, D. R. (1977). Some optimal designs for sampling in two dimensions, *Biometrika* **64**: 605–611.

Breidt, F. J. (1995). Markov chain designs for one-per-stratum spatial sampling, *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, DC, pp. 356–361.

Dunn, R. and Harrison, A. R. (1993). Two-dimensional systematic sampling of land use, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **42**: 585–601.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *The Annals of Statistics* **7**(1): 1–26.

Grafström, A., Lundström, N. L. P. and Schelin, L. (2012). Spatially balanced sampling through the pivotal method, *Biometrics* **68**: 514–520.

Gross, S. (1980). Median estimation in sample surveys, *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 181–184.

Hall, P. (1992). *The bootstrap and Edgeworth expansion*, Springer Science & Business Media, New YOrk.

Lahiri, S. N. and Zhu, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs, *The Annals of Statistics* **34**(4): 1774–1813.

Lister, A. J. and Scott, C. T. (2009). Use of space-filling curves to select sample locations in natural resource monitoring studies, *Environmental Monitoring and Assessment* **149**: 71–80.

Munholland, P. L. and Borkowski, J. J. (1996). Simple latin square sampling + 1: a spatial design using quadrats, *Biometrics* **52**: 125–136.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data, *Journal of the American Statistical Association* **83**(401): 231–241.

Shao, J. and Sitter, R. R. (1996). Bootstrap for imputed survey data, *Journal of the American Statistical Association* **91**(435): 1278–1288.

Sitter, R. R. (1992a). A resampling procedure for complex survey data, *Journal of the American Statistical Association* **87**(419): 755–765.

Sitter, R. R. (1992b). Comparing three bootstrap methods for survey data, *Canadian Journal of Statistics* **20**(2): 135–154.

Stevens, D. L. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources, *Journal of the American Statistical Association* **99**: 262–278.

CHAPTER 2. BOOTSTRAP INFERENCE FOR THE FINITE POPULATION TOTAL UNDER COMPLEX SAMPLING DESIGNS

Zhonglei Wang and Jae Kwang Kim

Abstract

Bootstrap is a useful tool for accessing uncertainty of estimates and for constructing confidence intervals, but it may provide erroneous inference in survey sampling if the sampling design is ignored. Most studies about bootstrap methods are developed under simple random sampling. In this paper, we propose a new bootstrap method applicable to simple random sampling as well as some complex sampling designs, including Poisson sampling and probability proportional to size sampling. A major difference between the proposed method and most existing ones is that the finite population is bootstrapped based on a multinomial distribution by incorporating the sampling weights, and an asymptotically pivotal statistic is used to make inference for the finite population total. Theoretical properties of the proposed method are investigated, and we show that the proposed method is second-order accurate using the Edgeworth expansion. Two simulation studies are conducted to compare the proposed method with the Wald-type method based on the asymptotic normality of the design-unbiased estimator. Simulation results show that the proposed method is more preferable since its coverage rate is better.

Key Words: Confidence interval; Edgeworth expansion; Multinomial distribution; Second-order accurate.

2.1 Introduction

Bootstrap, first proposed by Efron (1979), is a simulation-based approach for accessing uncertainty of estimates and for constructing confidence intervals. Bootstrap is popular because it is easy to implement, and has a better coverage property compared with the Wald-type method under certain conditions (Hall, 1992). However, the classical bootstrap method is not applicable under most sampling designs, where the independent and identical distribution assumption fails, and modifications of the bootstrap method have been proposed to handle the sampling design features.

Researches on bootstrap in survey sampling have been mainly focused on variance estimation. Rao and Wu (1988) discussed a rescaling bootstrap method for stratified sampling, and they showed that the bootstrap-t intervals are second-order accurate if the variance component is known. Sitter (1992a) considered a mirror-match bootstrap method for sampling designs without replacement. Sitter (1992b) extended the without-replacement bootstrap method to complex sampling designs, and compared the proposed method with the rescaling bootstrap method (Rao and Wu, 1988) and the mirror-match bootstrap method (Sitter, 1992a). Shao and Sitter (1996) proposed a bootstrap method for the case when survey data are subject to missingness. Beaumont and Patak (2012) proposed a generalized bootstrap method for Poisson sampling. Antal and Tillé (2011) proposed one-one resampling methods to estimate the variance for some complex sampling designs. Mashreghi et al. (2016) made a comprehensive overview of the bootstrap methods in survey sampling.

In this paper, we focus on another important usage of the bootstrap method: confidence interval estimation. We propose a new bootstrap method that can be applicable to some popular sampling designs, including Poisson sampling, simple random sampling (SRS) and the probability proportional to size (PPS) sampling. In the proposed method, a random mechanism is used to bootstrap the finite populations, from which the same sampling design is conducted to generate bootstrap samples. A similar idea has been successfully applied to SRS by Gross (1980) and Chao and Lo (1985), where the finite population is reconstructed by repeating the whole sample. Bickel and Freedman (1984) extended it to stratified sampling. Booth et al. (1994) proposed a bootstrap method to reconstruct the finite population using a different random mechanism from the one discussed by Bickel and Freedman (1984), and argued that the constructed confidence interval of a smooth function of the finite population mean is second-order accurate. Sverchkov and Pfeffermann (2004) proposed using a multinomial distribution to reconstruct the finite population to estimate the mean square error. However, up to our knowledge, there are no solid theoretical results of the bootstrap method for statistical inference under complex sampling designs.

The goal of this study is to develop a bootstrap method that approximates the sampling distribution of the design-unbiased estimator under popular sampling designs. The proposed method uses a multivariate binomial distribution to determine the number of copies for each sampled element by incorporating the sampling weights, and it differs from the one proposed by Sverchkov and Pfeffermann (2004) in the sense that the finite population is iteratively bootstrapped, and an asymptotically pivotal statistic is used to make inference for the finite population total. The proposed method is shown to be second-order accurate in the sense of having coverage error of order $o_p(n^{-1/2})$ (DiCiccio and Romano, 1995).

2.2 Sampling design and estimation

In survey sampling, the finite population is often assumed to be fixed, and the randomness is due to the sampling design. Let $\mathcal{F}_N = \{y_1, \ldots, y_N\}$ be the realized values of the study variable in the finite population of size N, and we are interested in making inference for the finite population total $Y = \sum_{i=1}^{N} y_i$ by survey data under three popular sampling designs, that is, Poisson sampling, SRS and PPS sampling. To avoid unnecessary details, we assume that the finite population size N is known, so it is equivalent to make inference for the finite population mean $\bar{Y} = N^{-1}Y$. Besides, we assume that the study variable is scalar.

For $i = 1, \ldots, N$, we denote the sample indicator of the i-th element to be I_i, which takes the value 1 if the i-th element is sampled and 0 otherwise. For Poisson sampling, a sample is created based on N independent Bernoulli trials, one for each element in the finite population. That is, $I_i \sim \mathrm{Ber}(\pi_i)$ for $i = 1, \ldots, N$, where $\mathrm{Ber}(\pi_i)$ is a Bernoulli distribution with a success probability

$\pi_i \in (0,1]$, and a sample is the collection of elements with $I_i = 1$. For without-replacement sampling designs, $\pi_i = E(I_i)$ is called the first-order inclusion probability of the i-th element, where the expectation is taken with respect to the sampling design. Let $\Pi_N = \{\pi_1, \ldots, \pi_N\}$ be the set of first-order inclusion probabilities for the finite population, and there is a one-to-one correspondence between y_i and π_i for $i = 1, \ldots, N$. Let $n = \sum_{i=1}^N I_i$ be a realized sample size, and $n_0 = E(n) = \sum_{i=1}^N \pi_i$ be its expectation under Poisson sampling. For SRS, a without-replacement sample of size n is selected with equal probabilities. The first-order inclusion probabilities are assumed to be available for the sampled elements, and $\pi_i = nN^{-1}$ under SRS. Denote $\hat{Y}_{Poi} = \sum_{i=1}^N I_i \pi_i^{-1} y_i$ to be the Horvitz–Thompson estimator (Horvitz and Thompson, 1952) of Y under Poisson sampling, and $\hat{Y}_{SRS}$ under SRS can be defined similarly. Without loss of generality, assume that the first n elements are sampled under Poisson sampling and SRS, and the variance estimates are denoted as $\hat{V}_{Poi} = \sum_{i=1}^n y_i^2 \pi_i^{-2}(1-\pi_i)$ and $\hat{V}_{SRS} = N(N-n)n^{-1}s_{SRS}^2$, respectively, where s_{SRS}^2 is the sample variance of $\{y_1, \ldots, y_n\}$. We assume that there is at least one non-zero element y_i in the finite population.

PPS sampling generates a sample of size n by independently selecting a single element from the same finite population for n times. Let $p_i \in (0,1)$ be the selection probability of the element y_i for $i = 1, \ldots, N$ with $\sum_{i=1}^N p_i = 1$. Denote $\mathcal{P}_N = \{p_1, \ldots, p_N\}$ to be the set of selection probabilities of $\mathcal{F}_N$, and there is a one-to-one correspondence between y_i and p_i for $i = 1, \ldots, N$. Assume that the selection probabilities of the sample are known. Let a_i be the index of the selected element for the i-th draw. We use the Hansen–Hurwitz estimator (Hansen and Hurwitz, 1943) to estimate Y, and denote it as $\hat{Y}_{PPS} = n^{-1}\sum_{i=1}^n z_i$, where $z_i = p_{a,i}^{-1} y_{a,i}$, and $p_{a,i} = p_k$ and $y_{a,i} = y_k$ if $a_i = k$. The variance estimator is $\hat{V}_{PPS} = n^{-1}(n-1)^{-1}\sum_{i=1}^n (z_i - \bar{z}_n)^2$ with $\bar{z}_n = n^{-1}\sum_{i=1}^n z_i$. For PPS sampling, we assume that there exist at least two distinguished values in $\{p_i^{-1} y_i : i = 1, \ldots, N\}$.

2.3 Bootstrap method for Poisson sampling

We propose to use the following bootstrap method to approximate the sampling distribution of $T_{Poi} = \hat{V}_{Poi}^{-1/2}(\hat{Y}_{Poi} - Y)$ under Poisson sampling.

Step 1. Based on the current sample of size n, generate $(N_1^*, \ldots, N_n^*)$ by a multinomial distribution $\mathrm{MN}(N;\rho)$ with N trials and a probability vector ρ, where $\rho = (\rho_1, \cdots, \rho_n)$ and $\rho_i = (\sum_{j=1}^n \pi_j^{-1})^{-1}\pi_i^{-1}$ for $i = 1, \ldots, n$. Denote $\mathcal{F}_N^* = \{y_1^*, \ldots, y_N^*\}$ and $\Pi_N^* = \{\pi_1^*, \ldots, \pi_N^*\}$, and they consist of N_i^* copies of y_i and π_i, respectively. Let the bootstrap finite population total be $Y^* = \sum_{i=1}^N y_i^* = \sum_{i=1}^n N_i^* y_i$.

Step 2. For each $i = 1, \cdots, n$, generate m_i^* independently by a binomial distribution $\mathrm{Bin}(N_i^*, \pi_i)$ with N_i^* trials and a success probability π_i. The bootstrap sample consists of m_i^* replicates of y_i under Poisson sampling. Denote $\hat{Y}_{Poi}^* = \sum_{i=1}^n m_i^* \pi_i^{-1} y_i$, and we can obtain $T_{Poi}^* = (\hat{V}_{Poi}^*)^{-1/2}(\hat{Y}_{Poi}^* - Y^*)$, where $\hat{V}_{Poi}^*$ is the bootstrap variance estimator obtained by applying the bootstrap sample to $\hat{V}_{Poi}$.

Step 3. Repeat the two steps above independently for M times.

Step 1 corresponds to generating the bootstrap finite population and the first-order inclusion probabilities. Step 2 is used to generate a bootstrap sample under Poisson sampling based on the bootstrap finite population, and a bootstrap replicate of T_{Poi} is obtained. Instead of using the Wald-type method, we propose to use T_{Poi}^* to make inference for T_{Poi}. That is, we use the quantiles of T_{Poi}^* to construct a confidence interval for T_{Poi}, which provides a confidence interval for Y. The proposed method would not work when the elements of the bootstrap sample are all zero, and this case is so trivial that we omit it if it occurs.

Before discussing the theoretical properties of the proposed method, we introduce some conditions on $\mathcal{F}_N$ and Π_N. For simplicity, we implicitly assume that y_i and π_i are indexed by N.

(C1) There exists a constant $\alpha \in (2^{-1}, 1]$ such that $n_0 = O(N^\alpha)$, and π_i satisfies $C_1 \le N{n_0}^{-1}\pi_i \le C_2$ for $i = 1, \ldots, N$, where C_1 and C_2 are positive constants with respect to N.

(C2) The eighth finite population moment satisfies $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N y_i^8 = C_3$, where C_3 is positive constants with respect to N.

(C3) The elements in the finite population satisfy $\max_{1\le i\le N} y_i = o(N^{(1-\alpha)/6})$.

We briefly comment on these conditions. The first part of (C1) is a mild restriction on the expected sample size, and the second part regulates the first-order inclusion probabilities. Condition (C1) is commonly used in survey sampling (Fuller, 2009). The moment condition mentioned in (C2) is used to guarantee the convergence of estimates of the variance and other finite population quantities, and it is also required for SRS and PPS sampling that we would discuss in the following sections. Condition (C2) can be released if we require the eighth finite population moment to be $O(1)$. Condition (C3) is used to show the consistency of the variance and the third central moment based on the bootstrap finite population.

Denote $(\mathcal{F}_N, \mathcal{B}_N, P_{N,Poi})$ to be a probability space, where $\mathcal{B}_N$ is the power set of $\mathcal{F}_N$, and $P_{N,Poi}(\cdot)$ is a probability measure on $\mathcal{F}_N$ associated with Poisson sampling. That is, $P_{N,Poi}(I_i = 1) = \pi_i$, and I_i and I_j are independent if $i \neq j$. For any positive integer set $J \subset \mathbb{N}$, denote $\mathbb{P}_{J,Poi} = \bigotimes_{j\in J} P_{j,Poi}$ to be the product probability measure on the product space $\bigotimes_{j\in J} \mathcal{F}_j$, where $\mathbb{N}$ denotes the set of positive integers; see §5.1 in Athreya and Lahiri (2006) for details. It can be shown that $\{\mathbb{P}_{J,Poi} : J \in \mathbb{N}\}$ is a consistent family of finite-dimensional distributions (Klenke, 2014, §14.3). Thus, by the Kolmogorov's consistency theorem, there exists a probability measure $\mathbb{P}_{Poi}$ on $U = \bigotimes_{N=1}^{\infty} \mathcal{F}_N$ equipped with the product σ-algebra $\mathcal{B}$, such that $\mathbb{P}_{J,Poi} = \mathbb{P}_{Poi} \circ \xi_J^{-1}$ for all finite positive integer set $J \subset \mathbb{N}$, where ξ_J is the canonical projection from U to $\bigotimes_{j\in J} \mathcal{F}_j$ (Klenke, 2014, §14.1).

Lemma 2.1. *Suppose that (C1) to (C2) hold. Then, we have*

$$\frac{n_0}{N^2} V_{Poi} = O(1), \tag{2.1}$$

$$\frac{n_0^2}{N^3} \mu_{Poi}^{(3)} = O(1), \tag{2.2}$$

where $V_{Poi} = \sum_{i=1}^N \pi_i^{-1}(1-\pi_i)y_i^2$ *and* $\mu_{Poi}^{(3)} = \sum_{i=1}^N y_i^3(1-\pi_i)\{(1-\pi_i)^2\pi_i^{-2} - 1\}$. *Furthermore,*

$$\frac{n_0}{N^2}(\hat{V}_{Poi} - V_{Poi}) \to 0 \tag{2.3}$$

as $N \to \infty$ *almost surely* $(\mathbb{P}_{Poi})$. *Besides,*

$$\frac{n_0^2}{N^3}(\hat{\mu}_{Poi}^{(3)} - \mu_{Poi}^{(3)}) = o_p(1), \tag{2.4}$$

where $\hat{\mu}^{(3)}_{Poi} = \sum_{i=1}^{n} \pi_i^{-1} y_i^3 (1-\pi_i)\{(1-\pi_i)^2 \pi_i^{-2} - 1\}$.

Lemma 2.1 shows that the ratio $\hat{V}_{Poi}^{-1} V_{Poi} \to 1$ almost surely, and $(\hat{\mu}^{(3)}_{Poi})^{-1} \mu^{(3)}_{Poi} \to 1$ in probability. Both results are essential for the theoretical property of the proposed bootstrap method under Poisson sampling.

Theorem 2.2. *Suppose that (C1) to (C2) hold. Then, we have*

$$\frac{\hat{\mu}^{(3)}_{Poi}}{\hat{V}^{3/2}_{Poi}} = O_p(n_0^{-1/2}). \tag{2.5}$$

Furthermore,

$$\hat{F}_{Poi}(z) = \Phi(z) + \frac{\hat{\mu}^{(3)}_{Poi}}{6\hat{V}^{3/2}_{Poi}}(1-z^2)\phi(z) + o_p(n_0^{-1/2}) \tag{2.6}$$

almost surely $(\mathbb{P}_{Poi})$ *for* $z \in \mathbb{R}$, *where* $\hat{F}_{Poi}(z)$ *is the cumulative distribution function of* T_{Poi} *under Poisson sampling,* $\Phi(z)$ *is the cumulative distribution function of the standard normal distribution with the density function* $\phi(z)$.

Theorem 2.2 shows the Edgeworth expansion for the distribution of the asymptotically pivotal statistic T_{Poi}. The Wald-type method, which is based on the asymptotic normality of T_{Poi}, is not second-order accurate, and its inference may be largely biased if the sample size is small and the value of $\hat{\mu}^{(3)}_{Poi}$ is not negligible.

Theorem 2.3. *Suppose that (C1) to (C3) hold. Then, we have*

$$\hat{F}^*_{Poi}(z) = \Phi(z) + \frac{\hat{\mu}^{(3)}_{Poi}}{6\hat{V}^{3/2}_{Poi}}(1-z^2)\phi(z) + o_p(n_0^{-1/2}) \tag{2.7}$$

almost surely conditional on the sample $\{y_1, \ldots, y_n\}$ *obtained by Poisson sampling in probability for* $z \in \mathbb{R}$, *where* $\hat{F}^*_{Poi}(z)$ *is the cumulative distribution function of* T^*_{Poi} *conditional on the realized sample.*

Theorem 2.3 presents the Edgeworth expansion for the distribution of T^*_{Poi} based on the proposed method. By comparing (2.6) in Theorem 2.2 with (2.7) in Theorem 2.3, we show that the proposed method is second-order accurate, and, hence, more preferable compared with the Wald-type method.

2.4 Bootstrap method for SRS sampling

We propose to use the following procedure to make inference for $T_{SRS} = \hat{V}_{SRS}^{-1/2}(\hat{Y}_{SRS} - Y)$ under SRS.

Step 1. Generate $(N_1^*, \ldots, N_n^*)$ by $\mathrm{MN}(N; \rho)$, where $\rho_i = n^{-1}$ for $i = 1, \ldots, n$. Then, $\mathcal{F}_N^*$ contains N_i^* copies of y_i for $i = 1, \ldots, n$.

Step 2. Generate a bootstrap sample of size n from $\mathcal{F}_N^*$ using SRS. Then, we can obtain $T_{SRS}^* = (\hat{V}_{N,SRS}^*)^{-1/2}(\hat{Y}_{SRS}^* - Y^*)$, where $\hat{Y}_{SRS}^*$ and $\hat{V}_{N,SRS}^*$ are the bootstrap versions of the $\hat{Y}_{SRS}$ and its variance estimator $\hat{V}_{N,SRS}$.

Step 3. Repeat the two steps above independently for M times.

The three steps for SRS are similar to those for Poisson sampling, but we do not construct Π_N^* since $\pi_i^* = nN^{-1}$ for $i = 1, \ldots, N$. We list some necessary conditions for studying the theoretical properties of the proposed method under SRS.

(C4) There exist $\beta \in (2^{-1}, 1]$ and $\kappa \in (0, 1)$ such that $n = O(N^\beta)$ and $nN^{-1} \le 1 - \kappa$ as $N \to \infty$.

(C5) The distribution $G_N(\cdot)$ converges weakly to a strongly non-lattice distribution $G(\cdot)$, where $G_N(x) = N^{-1}\sum_{i=1}^N \delta_{\{y_i\}}(x)$ and $\delta_A(x) = 1$ if $x \in A$ and 0 otherwise for a set A.

Condition (C4) is a counterpart of (C1), and it is used to rule out the trivial case when the sample size equals to that of the finite population. Condition (C5) is used to make the discussion easier, and a distribution $G(x)$ is strongly non-latticed if $|\int \exp(itx)\mathrm{d}G(x)| \neq 1$ for all $t \neq 0$; see Babu and Singh (1984) for more details. We would remark on the lattice case at the end of this section.

We can use a similar argument made in §2.3 to show that there exists a probability measure $\mathbb{P}_{SRS}$ on $U = \bigotimes_{N=1}^\infty \mathcal{F}_N$ equipped with the product σ-algebra $\mathcal{B}$ under SRS.

Lemma 2.4. *Suppose that (C2) and (C4) hold. Then,*

$$\sigma^2_{SRS} = O(1), \tag{2.8}$$

$$\mu^{(3)}_{SRS} = O(1), \tag{2.9}$$

where σ^2_{SRS} *and* $\mu^{(3)}_{SRS}$ *are the variance and the third central moment of* $\mathcal{F}_N$ *with respect to the distribution* $G_N(\cdot)$*. Besides,*

$$s^2_{SRS} - \sigma^2_{SRS} \to 0 \tag{2.10}$$

almost surely $(\mathbb{P}_{SRS})$ *as* $N \to \infty$*. Furthermore,*

$$\hat{\mu}^{(3)}_{SRS} - \mu^{(3)}_{SRS} = o_p(1), \tag{2.11}$$

where $\hat{\mu}^{(3)}_{SRS} = n^{-1}\sum_{i=1}^n y_i^3 + 2\bar{y}_n^3 - 3\bar{y}_n n^{-1}\sum_{i=}^n y_i^2$*, and* $\bar{y}_n = n^{-1}\sum_{i=1}^n y_i$ *is the sample mean.*

Lemma 2.4 shows the convergence properties of the sample variance and third central moment under mild conditions.

Theorem 2.5. *Suppose that (C2), (C4) and (C5) hold. Then, we have*

$$\hat{F}_{SRS}(z) = \Phi(z) + \frac{(1-2nN^{-1})\hat{\mu}^{(3)}_{SRS}}{6\{n(1-nN^{-1})\}^{1/2}s^3_{SRS}}(1-z^2)\phi(z) + o_p(n^{-1/2}) \tag{2.12}$$

almost surely $(\mathbb{P}_{SRS})$ *for* $z \in \mathbb{R}$*, where* $\hat{F}_{SRS}(z)$ *is the cumulative distribution function of* T_{SRS} *under SRS.*

Theorem 2.5 shows the Edgeworth expansion for the distribution of T_{SRS}. The bias of the Wald-type method is $O(n^{1/2})$ if $\mu^{(3)}_{SRS} \neq 0$. We can use a similar procedure discussed by Babu and Singh (1985) to prove Theorem 2.5 based on Lemma 2.4, and we omit the proof in this paper.

Theorem 2.6. *Suppose that (C2), (C4) and (C5) hold. Then, we have*

$$\hat{F}^*_{SRS}(z) = \Phi(z) + \frac{(1-2nN^{-1})\hat{\mu}^{(3)}_{SRS}}{6\{n(1-nN^{-1})\}^{1/2}s^3_{SRS}}(1-z^2)\phi(z) + o_p(n^{-1/2}) \tag{2.13}$$

almost surely conditional on the sample $\{y_1, \ldots, y_n\}$ *obtained by SRS in probability for* $z \in \mathbb{R}$*, where* $\hat{F}^*_{SRS}(z)$ *is the cumulative distribution function of* T^*_{SRS} *conditional on the realized sample.*

Theorem 2.6 shows the Edgeworth expansion for the distribution of T^*_{SRS} based on the proposed method. By comparing (2.12) with (2.13), we show the second-order accuracy of the proposed method.

Remark 1. Based Lemma 2.4 and the proof of Theorem 2.6, we can show that the proposed method is still second-order accurate if the limiting distribution $G(\cdot)$ is latticed under certain conditions (Babu and Singh, 1985) by noting that the function $h(x) = [x] - x + 2^{-1}$ is continuous almost everywhere with respect to the Lebesgue measure. We omit the proof for the latticed case in this paper.

2.5 Bootstrap method for PPS sampling

We consider PPS sampling in this section, and propose to use the following bootstrap method to approximate the sampling distribution of $T_{PPS} = \hat{V}_{PPS}^{-1/2}(\hat{Y}_{PPS} - Y)$.

Step 1. Obtain $(N^*_{a,1}, \ldots, N^*_{a,n})$ by a multinomial distribution $\mathrm{MN}(N; \rho)$, where $\rho_{a,i} = p_{a,i}^{-1}(\sum_{j=1}^n p_{a,j}^{-1})^{-1}$ for $i = 1, \ldots, n$. Then, $\mathcal{F}^*_N$ consists of $N^*_{a,i}$ copies of $y_{a,i}$, and the bootstrap finite population total is denoted as $Y^* = \sum_{i=1}^n N^*_{a,i} y_{a,i}$. The set of bootstrap selection probabilities is denoted as $\mathcal{P}^*_N = \{(C^*_N)^{-1}p^*_1, \ldots, (C^*_N)^{-1}p^*_N\}$, where $C^*_N = \sum_{i=1}^N p^*_i = \sum_{i=1}^n N^*_{a,i} p_{a,i}$, and $\{p^*_1, \ldots, p^*_N\}$ consists of $N^*_{a,i}$ copies of $p_{a,i}$ for $i = 1, \ldots, n$.

Step 2. Based on $\mathcal{F}^*_N$, sample one element with selection probability $(C^*_N)^{-1}p^*_i$ for the i-th element independently for n times. Then, we have $\hat{Y}^*_{PPS} = \sum_{i=1}^n n^{-1} C^*_N (p^*_{b,i})^{-1} y^*_{b,i}$ and $T^*_{PPS} = (\hat{V}^*_{N,PPS})^{-1/2}(\hat{Y}^*_{PPS} - Y^*)$, where $y^*_{b,i} = y^*_k$ and $p^*_{b,i} = p^*_k$ if the index of the i-th draw is k, and $\hat{V}^*_{N,PPS}$ is the counterpart of $\hat{V}_{N,PPS}$ based on the bootstrap sample.

Step 3. Repeat the two steps above independently for M times.

The proposed method under PPS sampling is a counterpart of the one for Poisson sampling. Before drawing a sample, the bootstrap selection probability should be standardized. It is trivial when $(p^*_{b,i})^{-1} y^*_{b,i}$ is the same for $i = 1, \ldots, n$, and we omit it if it occurs.

Similarly to §2.3, we propose to use T^*_{PPS} to make inference for T_{PPS}. The following regular condition is required to validate the proposed method under PPS sampling.

(C6) There exists $\gamma \in (2^{-1}, 1]$ such that $n = O(N^{\gamma})$, the selection probability satisfies

$$C_4 \leq Np_i \leq C_5$$

for $i = 1, \ldots, N$, where C_4 and C_5 are positive constants with respect to N.

Condition (C6) regulates the sample size and the selection probability.

Based on a similar argument made under Poisson sampling, there exists a probability measure $\mathbb{P}_{PPS}$ on $U = \bigotimes_{N=1}^{\infty} \mathcal{F}_N$ equipped with the product σ-algebra $\mathcal{B}$.

Lemma 2.7. *Suppose that (C2) and (C6) hold. Then,*

$$N^{-2}\sigma^2_{PPS} = O(1), \tag{2.14}$$

$$N^{-3}\mu^{(3)}_{PPS} = O(1), \tag{2.15}$$

where and $\sigma^2_{PPS} = \sum_{i=1}^{N} p_i(p_i^{-1}y_i - Y)^2$ *and* $\mu^{(3)}_{PPS} = \sum_{i=1}^{N} p_i(p_i^{-1}y_i - Y)^3$. *Furthermore,*

$$N^{-2}(s^2_{PPS} - \sigma^2_{PPS}) \to 0 \tag{2.16}$$

almost surely $(\mathbb{P}_{PPS})$ *as* $N \to \infty$, *where* s^2_{PPS} *is the sample variance of* $\{z_i : i = 1, \ldots, n\}$. *Besides,*

$$N^{-3}(\hat{\mu}^{(3)}_{PPS} - \mu^{(3)}_{PPS}) = o_p(1), \tag{2.17}$$

where $\hat{\mu}^{(3)}_{PPS} = n^{-1}\sum_{i=1}^{n} z_i^3 + 2\bar{z}_n^3 - 3\bar{z}_n n^{-1}\sum_{i=}^{n} z_i^2$.

Lemma 2.7 shows similar results as those in Lemma 2.1 under PPS sampling. In Lemma 2.7, we only list some key results used for the Edgeworth expansion of T_{PPS}. More similar results can be derived, and we would discuss them when proving Theorem 2.9.

Theorem 2.8. *Suppose that (C2) and (C6) hold. Then, we have*

$$\hat{F}_{PPS}(z) = \Phi(z) + \frac{\hat{\mu}^{(3)}_{PPS}}{6\sqrt{n}s^3_{PPS}}(1 - z^2)\phi(z) + o_p(n^{-1/2}) \tag{2.18}$$

almost surely $(\mathbb{P}_{PPS})$ *as* $N \to \infty$, *where* $\hat{F}_{PPS}$ *is the cumulative distribution function of* T_{PPS} *under PPS sampling.*

Theorem 2.8 shows the Edgeworth expansion for the distribution of T_{PPS}. Based on the result in this theorem, the Wald-type method may provide inefficient inference results for Y if the sample size is small and the value of $\hat{\mu}_{PPS}^{(3)}$ is not ignorable.

Theorem 2.9. *Suppose that (C2) and (C6) hold. Then, we have*

$$\hat{F}^*_{PPS}(z) = \Phi(z) + \frac{\hat{\mu}_{PPS}^{(3)}}{6\sqrt{n}s_{PPS}^3}(1 - z^2)\phi(z) + o_p(n^{-1/2}) \tag{2.19}$$

almost surely as $N \to \infty$ conditional on the sample in probability, where $\hat{F}^*_{PPS}(z)$ is the conditional distribution of T^*_{PPS} under the realized sample.

Theorem 2.9 shows the Edgeworth expandion for the cumulative distribution function of T^*_{PPS} based on the proposed method. By comparing (2.18) with (2.19), we have shown that the proposed method is second-order accurate, and can produce more reliable inference for Y than the Wald-type method under PPS sampling.

2.6 Simulation study

2.6.1 Single-stage sampling designs

We conduct a simulation study based on single-stage sampling designs in this section. A finite population $\mathcal{F}_N = \{y_1, \ldots, y_N\}$ is generated by

$$y_i \sim \text{Ex}(10)$$

for $i = 1, \ldots, N$, where $\text{Ex}(\lambda)$ is an exponential distribution with a scale parameter λ, and the finite population size is $N = 500$. The size measure is simulated by $z_i = \log(3 + s_i)$ for $i = 1, \ldots, N$, where $s_i \mid y_i \sim \text{Exponential}(y_i)$. The expected sample size is $n_0 \in \{10, 100\}$, and we assume that the finite population size N is known. We are interested in constructing a 90% confidence interval for the finite population mean $\bar{Y}$ by survey data under the following sampling designs, and its true value is around 9.7 in this simulation.

1. Poisson sampling. The first-order inclusion probability is $\pi_i = n_0 z_i \left(\sum_{j=1}^{N} z_j\right)^{-1}$ for $i = 1, \ldots, N$.

2. SRS with sample size n_0.

3. PPS sampling. The selection probability for this design is $p_i = z_i \left(\sum_{j=1}^{N} z_j\right)^{-1}$ for $i = 1, \ldots, N$, and the sample size is n_0.

For a sample, denote $\tilde{V}$ to be the variance estimator for $\tilde{Y}$, where $\tilde{Y}$ is the design-unbiased estimate of $\bar{Y}$ under a specific sampling design. We consider the following methods to construct the 90% confidence interval.

Method I. Proposed bootstrap method by setting $M = 1,000$. Denote $q_{B,0.05}$ and $q_{B,0.95}$ to be the 5%-th and 95%-th quantiles of the bootstrap quantities $\{(\tilde{V}^{*(m)})^{-1/2}(\tilde{Y}^{*(m)} - \bar{Y}^{*(m)}) : m = 1, \ldots, M\}$ obtained by the proposed method, where $\tilde{V}^{*(m)}$, $\tilde{Y}^{*(m)}$ and $\bar{Y}^{*(m)}$ are the bootstrap counterparts of $\tilde{V}$, $\tilde{Y}$ and $\bar{Y}$ in the m-th repetition. Then, a 90% confidence interval can be constructed by

$$(\tilde{Y} - q_{B,0.95}\tilde{V}^{1/2}, \tilde{Y} - q_{B,0.05}\tilde{V}^{1/2}).$$

Method II. Wald-type method. A 90% confidence interval is constructed by

$$(\tilde{Y} - q_{0.95}\tilde{V}^{1/2}, \tilde{Y} - q_{0.05}\tilde{V}^{1/2}),$$

where $q_{0.05}$ and $q_{0.95}$ are the 5%-th and 95%-th quantiles of the standard normal distribution.

We conduct $1,000$ Monte Carlo simulations for each sampling design, and the two methods are compared in terms of the coverage rate and the length of the constructed 90% confidence interval. Table 2.1 summarizes the simulation results. When the sample size is small, the proposed method is more preferable in the sense that its coverage rates are closer to 0.9 compared with the Wald-type method under the three sampling designs. The confidence interval constructed by the proposed method is more conservative. As the sample size increases to $n_0 = 100$, the performance of the two methods is approximately the same under SRS and PPS sampling in the sense that the coverage rates of both methods are close to 0.9, and confidence interval lengths are approximately the same. Under Poisson sampling, the coverage rate of the proposed method is 0.9 when $n_0 = 100$, but that of the Wald-type method is only 0.87, showing the poor performance even when n_0 is reasonably large.

Table 2.1: Coverage rate and length of the constructed 90% confidence interval for the proposed method (Bootstrap) and the Wald-type method (Wald-type) under Poisson sampling (Poisson), SRS and PPS sampling (PPS). “C.R.” stands for the coverage rate, and “C.L.” presents the Monte Carlo mean of the lengths of the confidence interval.

Design	Method	$n_0 = 10$		$n_0 = 100$	
		C.R.	C.L.	C.R.	C.L.
Poisson	Bootstrap	0.89	15.2	0.89	3.6
	Wald-type	0.83	11.9	0.87	3.5
SRS	Bootstrap	0.86	11.3	0.90	2.8
	Wald-type	0.82	8.9	0.90	2.8
PPS	Bootstrap	0.88	10.3	0.90	2.6
	Wald-type	0.82	7.5	0.89	2.5

2.6.2 Two-stage sampling designs

In this section, we check the performance of the proposed method under two-stage sampling designs. A finite population $\mathcal{F}_N = \{y_{i,j} : i = 1, \ldots, H; j = 1, \ldots, N_i\}$ is generated by

$$
\begin{aligned}
y_{i,j} &= 50 + a_i + e_{i,j}, \\
a_i &\sim N(0, 50), \\
e_{i,j} &\sim \mathrm{Ex}(20), \\
N_i \mid a_i &\sim \mathrm{Poisson}(q_i) + c_0,
\end{aligned}
$$

for $i = 1, \ldots, H$ and $j = 1, \ldots, N_i$, where Poisson(λ) is a Poisson distribution with a rate parameter λ, $q_i = (a_i - 25)^2/20$, $c_0 = 40$ is the minimum cluster size, and $H = 100$ is the number of clusters in the finite population. The finite population size is $N = 17,011$, and the cluster sizes range from 40 to 542. We assume that the finite population size N and cluster sizes $N_1, \ldots, N_H$ are known. We are interested in constructing 90% confidence intervals for the finite population mean $\bar{Y} = N^{-1}\sum_{i=1}^{H}\sum_{j=1}^{N_i} y_{i,j}$ and the finite population proportion $P = N^{-1}\sum_{i=1}^{H}\sum_{j=1}^{N_i}\delta_{(-\infty,q_y)}(y_{i,j})$, where q_y is the 50%-th quantile of the finite population $\mathcal{F}_N$.The true value of $\bar{Y}$ is approximately 69.1.

We consider two different sampling designs for the first stage; one is Poisson sampling, and the other one is PPS sampling. The first-order inclusion probability (selection probability) of the i-th cluster is proportional to its cluster size N_i under Poisson (PPS) sampling for $i = 1, \dots, H$. SRS is conducted within each selected cluster independently in the second stage. The expected sample size of the first stage is n_1, and the one of the second stage is n_2. In this simulation, we consider $n_1 \in \{10, 30\}$ and $n_2 \in \{10, 30\}$.

The derivations of the design-unbiased estimator $\tilde{Y}$ and its variance estimator $\tilde{V}$ are presented in §2.8.2, where $\tilde{Y}$ and $\tilde{V}$ are defined in §2.6.1. Similar results hold for the corresponding estimates for P. We use the following methods to construct the 90% confidence intervals for the parameters we are interested in.

Method I. The proposed method extended to a two-level stage sampling design. This method is approximately the same as the one mentioned in §2.6.1, but we set the repetition number M to be 500 and use the following two steps to bootstrap the finite population.

Step 1. Use the proposed method to bootstrap the H clusters by treating them as "elements", and the original sample within each selected cluster are replicated accordingly.

Step 2. For each bootstrap cluster, apply the proposed method to bootstrap the cluster finite population independently.

Method II. Wald-type method, which is the same as the one discussed in §2.6.1.

We conduct 500 Monte Carlo simulations for each finite population parameter. Table 2.2 summarizes the coverage rate and average length of the constructed 90% confidence interval for the estimate of the finite population mean. When sample size is small, the coverage rates of the proposed method are closer to 0.9 compared with its alternative under both two-stage sampling designs generally. The Wald-type method performs well in the case where Poisson sampling is conducted in the first stage, and its coverage rate is around 0.9 even when the sample size n_1 is small. However, the coverage rate of the Wald-type method is only 0.87 and 0.86 for the cases $n_2 = 10$ and $n_2 = 30$ when PPS sampling is used in the first stage and $n_1 = 10$, and they underestimate their nominal

Table 2.2: Coverage rate and length of the 90% confidence interval for $\bar{Y}$ by the proposed method (Bootstrap) and the Wald-type method (Wald-type) under two-stage sampling designs. For the sampling designs of the first stage, we consider Poisson sampling (Poisson) and PPS sampling (PPS), and SRS is used in the second stage. "C.R." shows the coverage rate, and "C.L." presents the Monte Carlo mean of the length for the 90% confidence interval.

Design		Method	$n_1 = 10$		$n_1 = 30$	
			C.R.	C.L.	C.R.	C.L.
Poisson	$n_2 = 10$	Bootstrap	0.91	74.8	0.90	34.4
		Wald-type	0.89	69.6	0.90	33.9
	$n_2 = 30$	Bootstrap	0.90	74.5	0.90	34.2
		Wald-type	0.89	69.4	0.90	33.6
PPS	$n_2 = 10$	Bootstrap	0.90	9.1	0.91	4.8
		Wald-type	0.87	8.0	0.90	4.7
	$n_2 = 30$	Bootstrap	0.90	7.8	0.89	4.1
		Wald-type	0.86	6.8	0.88	4.0

truth 0.9. The confidence intervals of the proposed method is more conservative than the ones of the Wald-type method when sample size is small. The proposed method has a similar performance with the Wald-type method when the sample size is large in the sense that their coverage rates and confidence interval lengths are approximately the same.

Table 2.3 shows the summary statistics for estimating the finite population proportion P. For each case, the coverage rate of the proposed method is around 0.9, and it is closer to its nominal truth compared with the one by the Wald-type method. When sample size is small, the performance of the Wald-type method is not good in the sense that the coverage rates are below 0.9 when $n_1 = 10$. Furthermore, the coverage rate of the Wald-type method is only 0.84 and 0.85 for the cases $n_2 = 10$ and $n_2 = 30$ when PPS sampling is used in the first stage and $n_1 = 10$. The constructed confidence interval by the proposed method is wider than the one by the Wald-type method when sample size is small. As n_1 increases to 30, the performance of the proposed method is approximately the same with the Wald-type method.

Table 2.3: Coverage rate and length of the 90% confidence interval for P by the proposed method (Bootstrap) and the Wald-type method (Wald-type) under two-stage sampling designs. For the sampling designs of the first stage, we consider Poisson sampling (Poisson) and PPS sampling (PPS), and the SRS design is used in the second stage. "C.R." shows the coverage rate, and "C.L." presents the Monte Carlo mean of the length for the 90% confidence interval.

Design		Method	$n_1 = 10$		$n_1 = 30$	
			C.R.	C.L.	C.R.	C.L.
Poisson	$n_2 = 10$	Bootstrap	0.91	0.6	0.89	0.3
		Wald-type	0.87	0.5	0.88	0.3
	$n_2 = 30$	Bootstrap	0.88	0.6	0.90	0.2
		Wald-type	0.87	0.5	0.91	0.2
PPS	$n_2 = 10$	Bootstrap	0.89	0.3	0.90	0.2
		Wald-type	0.84	0.2	0.90	0.2
	$n_2 = 30$	Bootstrap	0.90	0.3	0.90	0.1
		Wald-type	0.85	0.2	0.89	0.1

Remark 2. We have carefully examined the empirical distribution of $\tilde{Y}$, and it is approximately symmetric even when the sample size is small. This is the reason why both methods work well for constructing the 90% confidence interval of $\bar{Y}$. However, the empirical distribution of the proportion estimator is slightly right-skewed when $n_1 = 10$.

We have also done simulations to compare the proposed method with the nonparametric Bayesian bootstrap method discussed by Dong et al. (2014) and the one based on the two-step inverse sampling method proposed by Sverchkov and Pfeffermann (2004) under the single-stage and two-stage sampling designs. Simulation results show that the proposed method outperforms these two approaches in the sense that the coverage rate of the constructed confidence interval by the proposed method is much closer to the nominal truth when the sampling distribution of the design-unbiased estimator is skewed.

2.7 Conclusion

In this paper, we propose bootstrap methods for Poisson sampling, SRS and PPS sampling, and we show that the proposed method is second-order accurate. The first step of the propose method corresponds to an inverse sampling procedure, and the sampling weights are incorporated to bootstrap the finite population. Since the proposed method is based on an asymptotically pivotal statistic, it is necessary to estimate the variance of the design-based estimator. Simulation results show that the proposed method provides more conservative confidence interval than the Wald-type method when the sample size is small, and the constructed 90% confidence interval by the proposed has a better coverage rate. Although the proposed method is discussed under the single-stage sampling designs, simulation shows that it works well under two-stage sampling designs. It may be extended to other complex sampling designs when the asymptotic distribution of the design-unbiased estimator exists, but the second-order accuracy may not be guaranteed. Besides, the proposed method can be easily parallelized in practice.

2.8 Appendix

2.8.1 Proofs

For the clearance purpose, we explicitly express $y_{N,i}$, Y_N, $I_{N,i}$, $\pi_{N,i}$ and $p_{N,i}$ for y_i, Y, I_i, π_i and p_i to highlight that they are indexed by N, and the same notation is used for other quantities without mentioning explicitly. Denote $E(\cdot \mid \mathcal{F}_N)$ and $\text{var}(\cdot \mid \mathcal{F}_N)$ to be the expectation and variance with respect to the probability measure P_N under a specific sampling design, $E_*(\cdot)$ and $\text{var}_*(\cdot)$ to be the conditional mean and variance operators for the first steps of the proposed method conditional on the sample $\{y_{N,1}, \ldots, y_{N,n}\}$, and $E_{**}(\cdot)$ and $\text{var}_{**}(\cdot)$ to be the expectation and variance operator with respect to the sampling design conditional on the bootstrap finite population $\mathcal{F}_N^*$.

Proof of Lemma 2.1. Consider

$$
\begin{aligned}
n_0 N^{-2} \sum_{i=1}^{N} \pi_{N,i}^{-1} y_{N,i}^2 (1 - \pi_{N,i}) &\leq n_0 N^{-2} \sum_{i=1}^{N} \pi_{N,i}^{-1} y_{N,i}^2 \\
&= O(1),
\end{aligned}
$$

where the last equality holds by (C1) and (C2). Thus, we have proved (2.1). Similarly, we can prove (2.2) Based on some algebra and (C2), we have

$$\sum_{i\neq j} y_{N,i}^4 y_{N,j}^4 \quad = \quad O(N^2), \tag{2.20}$$

where $i \neq j$ stands for "$\{i,j\} \subset \{1,\ldots,N\}$ and $i \neq j$".

Denote $X_{N,i}^{(1)} = n_0 N^{-2}(I_{N,i}\pi_{N,i}^{-1} - 1) y_{N,i}^2 \pi_{N,i}^{-1}(1-\pi_{N,i})$, and $D_N^{(1)}$ to be the event $\{|\sum_{i=1}^N X_{N,i}^{(1)}| > \epsilon\}$ for $N \in \mathbb{N}$, where $\epsilon \in (0,\infty)$. By the Borel-Cantelli Lemma (Athreya and Lahiri, 2006, Thereom 7.2.2), to show (2.3), it is enough to prove

$$\sum_{N=1}^{\infty} \mathbb{P}_{Poi}(D_N^{(1)}) < \infty. \tag{2.21}$$

By the Markov's inequality (Athreya and Lahiri, 2006, Proposition 6.2.4), we have

$$\begin{aligned}
\mathbb{P}_{Poi}(D_N^{(1)}) &\leq \epsilon^{-4} E\left\{ \left(\sum_{i=1}^N X_{N,i}^{(1)} \right)^4 \mid \mathcal{F}_N \right\} \\
&= \epsilon^{-4} \left[\sum_{i=1}^N E\left\{ \left(X_{N,j}^{(1)} \right)^4 \mid \mathcal{F}_N \right\} + \sum_{i\neq j} E\left\{ \left(X_{N.i}^{(1)} \right)^2 \mid \mathcal{F}_N \right\} E\left\{ \left(X_{N,j}^{(1)} \right)^2 \mid \mathcal{F}_N \right\} \right],
\end{aligned}$$

where the last equality holds since $E(X_{N,i}^{(1)} \mid \mathcal{F}_N) = 0$ for $i = 1,\ldots,N$, and $X_{N,i}^{(1)}$ is independent of $X_{N,j}^{(1)}$ if $i \neq j$.

Consider

$$\begin{aligned}
E\left\{ \left(X_{N,i}^{(1)} \right)^4 \mid \mathcal{F}_N \right\} &= n_0^4 N^{-8} y_{N,i}^8 \pi_{N,i}^{-4} (1-\pi_{N,i})^4 \{\pi_{N,i}(\pi_{N,i}^{-1} - 1)^4 + (1-\pi_{N,i})\} \\
&\leq C_{N,1} n_0^{-3} N^{-1} y_{N,i}^8,
\end{aligned} \tag{2.22}$$

where $C_{N,1}$ is a constant determined by (C1). Next, consider

$$\begin{aligned}
E\left\{ \left(X_{N,i}^{(1)} \right)^2 \mid \mathcal{F}_N \right\} &= n_0^2 N^{-4} y_{N,i}^4 \pi_{N,i}^{-3} (1-\pi_{N,i})^3 \\
&\leq C_{N,2} n_0^{-1} N^{-1} y_{N,i}^4,
\end{aligned} \tag{2.23}$$

where $C_{N,2}$ is a constant. By (2.22) and (2.23), we have

$$\begin{aligned}
\mathbb{P}_{Poi}(D_N^{(1)}) &\leq \epsilon^{-4} C_{N,1} n_0^{-3} N^{-1} \sum_{i=1}^N y_{N,i}^8 + \epsilon^{-4} C_{N,2}^2 n_0^{-2} N^{-2} \sum_{i\neq j} y_{N,i}^4 y_{N,j}^4 \\
&\leq C_{N,3} n_0^{-3} + C_{N,4} n_0^{-2} \\
&= O(N^{-2\alpha}),
\end{aligned}$$

where the last inequality holds by (C2) and (2.20). Since $\alpha \in (2^{-1}, 1]$ by (C1), we have proved (2.3) based on (2.21).

Consider

$$\begin{aligned}
& n_0^2 N^{-3} E\left[\sum_{i=1}^{n} \pi_{N,i}^{-1} y_{N,i}^3 (1-\pi_{N,i})\{(1-\pi_{N,i})^2 \pi_{N,i}^{-2} - 1\} \mid \mathcal{F}_N\right] \\
& = n_0^2 N^{-3} \sum_{i=1}^{N} y_{N,i}^3 (1-\pi_{N,i})\{(1-\pi_{N,i})^2 \pi_{N,i}^{-2} - 1\} \qquad (2.24) \\
& \text{var}\left[n_0^2 N^{-3} \sum_{i=1}^{n} \pi_{N,i}^{-1} y_{N,i}^3 (1-\pi_{N,i})\{(1-\pi_{N,i})^2 \pi_{N,i}^{-2} - 1\} \mid \mathcal{F}_N\right] \\
& \leq n_0^4 N^{-6} \sum_{i=1}^{N} \pi_{N,i}^{-5} y_{N,i}^6 \\
= \; & O(n_0^{-1}) \qquad (2.25)
\end{aligned}$$

where the last equality of (2.25) by (C1) and (C2). Thus, by (2.24) and (2.25), we have proved (2.4) of Lemma 2.1. □

Proof of Theorem 2.2. Denote $F_{N,HT}(z) = P_{N,Poi}\{T_{N,Poi} < z\}$, where $T_{N,Poi} = V_{N,Poi}^{-1/2}(\hat{Y}_{N,HT} - Y_N)$. By the results in §16.6 discussed by Feller (2008), we have

$$F_{N,HT}(z) = \Phi(z) + \frac{\mu_{N,Poi}^{(3)}}{6V_{N,Poi}^{3/2}}(1-z^2)\phi(z) + R_N(z), \qquad (2.26)$$

where $R_N(z)$ is the remainder term.

By Lemma 2.1, we have

$$\begin{aligned}
\mu_{N,Poi}^{(3)} &= O(N^3 n_0^{-2}), \qquad (2.27) \\
V_{N,Poi} &= O(N^2 n_0^{-1}). \qquad (2.28)
\end{aligned}$$

Thus, by (2.27) and (2.28), we have

$$\frac{\mu_{N,Poi}^{(3)}}{V_{N,Poi}^{3/2}} = O(n_0^{-1/2}). \qquad (2.29)$$

Based on (2.29), (2.3) and (2.4) in Lemma 2.1, we have proved (2.5).

Denote $X_{N,i}^{(2)} = n_0^{1/2} N^{-1}(I_{N,i}\pi_{N,i}^{-1} - 1)y_{N,i}$, and it can be shown that $E(X_{N,i}^{(2)} \mid \mathcal{F}_N) = 0$. By a similar argument with (2.27), we have

$$\sum_{i=1}^{N} E\{(I_{N,i}\pi_{N,i}^{-1} - 1)^4 \mid \mathcal{F}_N\} y_{N,i}^4 \quad = \quad O(N^4 n_0^{-3}). \tag{2.30}$$

Based on (2.28) and (2.30), we have

$$\begin{aligned} & N^{1/2} V_{N,Poi}^{-3} \left[\sum_{i=1}^{N} E\{(I_{N,i}\pi_{N,i}^{-1} - 1)^4 \mid \mathcal{F}_N\} y_{N,i}^4 \right]^{3/2} \\ = \quad & O(n_0^{-1/2} N^{-\alpha+1/2}) \\ = \quad & o(n_0^{-1/2}), \end{aligned} \tag{2.31}$$

where the last equality holds by (C1). Thus, by (2.31) and the argument in §16.6 of Feller (2008), we have $R_N(z) = o(n_0^{-1/2})$ uniformly, and recall that $R_N(z)$ is the remainder term of (2.26). By Lemma 2.1, we have proved (2.6). □

Proof of Theorem 2.3. Without loss of generality, we use $I_{N,i}$ explicitly to denote the sample in this proof.

We first show

$$N^{-1} \sum_{i=1}^{N} (I_{N,i}\pi_{N,i}^{-1} - 1) \to 0 \tag{2.32}$$

almost surely ($\mathbb{P}_{Poi}$).

Denote $D_N^{(2)}$ to be the event $\{N^{-1} | \sum_{i=1}^{N} (I_{N,i}\pi_{N,i}^{-1} - 1) | > \epsilon\}$, where ϵ is a fixed positive number. Similar with the argument used in proving Lemma 2.1, consider

$$\begin{aligned} \mathbb{P}_{Poi}(D_N^{(2)}) \quad \leq \quad & \epsilon^{-4} N^{-4} E\left\{ \left| \sum_{i=1}^{N} (I_{N,i}\pi_{N,i}^{-1} - 1) \right|^4 \mid \mathcal{F}_N \right\} \\ = \quad & \epsilon^{-4} N^{-4} \left[\sum_{i=1}^{N} (1 - \pi_{N,i})\{(1 - \pi_{N,i})^3 \pi_{N,i}^{-3} + 1\} + \right. \\ & \left. \sum_{i \neq j} (1 - \pi_{N,i})\{(1 - \pi_{N,i})\pi_{N,i}^{-1} + 1\}(1 - \pi_{N,j})\{(1 - \pi_{N,j})\pi_{N,j}^{-1} + 1\} \right] \\ \leq \quad & C_{N,5}\epsilon^{-4} n_0^{-3} + C_{N,6}\epsilon^{-4} n_0^{-2}, \\ = \quad & O(n_0^{-2}), \end{aligned}$$

where $C_{N,5}$ and $C_{N,6}$ are constants with respect to N. Thus, we have $\sum_{N=1}^{\infty} \mathbb{P}_{Poi}(D_N^{(2)}) < \infty$, and we have proved (2.32) by the Borel-Cantelli Lemma.

Next, consider

$$\begin{aligned} E\left(n_0^{-1}\sum_{i=1}^{N} I_{N,i}y_{N,i}^8 \mid \mathcal{F}_N\right) &= n_0^{-1}\sum_{i=1}^{N} E(I_{N,i} \mid \mathcal{F}_N)y_{N,i}^8 \\ &= n_0^{-1}\sum_{i=1}^{N} \pi_{N,i}y_{N,i}^8 \\ &\leq C_1^{-1}N^{-1}\sum_{i=1}^{N} y_{N,i}^8 \\ &\to C_1^{-1}C_3, \end{aligned} \tag{2.33}$$

where the second inequality holds by (C1) and the convergence result holds by (C2). Thus, by (2.33) and the Markov's inequality, we have

$$n_0^{-1}\sum_{i=1}^{N} I_{N,i}y_{N,i}^8 = O_p(1). \tag{2.34}$$

Consider

$$E\left(n_0^{-1}\sum_{i=1}^{N} I_{N,i} \mid \mathcal{F}_N\right) = 1, \tag{2.35}$$

$$\begin{aligned} \operatorname{var}\left(n_0^{-1}\sum_{i=1}^{N} I_{N,i} \mid \mathcal{F}_N\right) &\leq n_0^{-2}\sum_{i=1}^{N} \pi_{N,i} \\ &\leq C_1^{-1}n_0^{-1}, \end{aligned} \tag{2.36}$$

where the second inequality of (2.36) holds by (C1). Thus, by (2.35) and (2.36), we have

$$n_0^{-1}\sum_{i=1}^{N} I_{N,i} = 1 + o_p(1). \tag{2.37}$$

Next, we consider

$$\begin{aligned} E_*\left\{N^{-1}\sum_{i=1}^{N} (y_{N,i}^*)^8\right\} &= N^{-1}\sum_{i=1}^{n} N\pi_{N,i}^{-1}\left(\sum_{i=1}^{n} \pi_{N,i}^{-1}\right)^{-1} y_{N,i}^8 \\ &\leq C_1^{-1}n_0^{-1}\sum_{i=1}^{N} I_{N,i}y_{N,i}^8 \\ &= O_p(1), \end{aligned} \tag{2.38}$$

where the first equality holds by the property of the proposed method, the inequality holds by (2.32) and (C1), and the last equality holds by (2.34). Thus, by (2.38) and Markov's inequality, we have

$$N^{-1}\sum_{i=1}^{N}(y^*_{N,i})^8 = O_p(1). \tag{2.39}$$

By a similar argument with (2.39) and Lemma 2.1, we have $\frac{n_0}{N^2}(\hat{V}^*_{N,Poi} - V^*_{N,Poi}) \to 0$ almost surely ($\mathbb{P}^*_{Poi}$) in probability, where $V^*_{N,Poi} = \sum_{i=1}^{N}(y^*_{N,i})^2(\pi^*_{N,i})^{-1}(1-\pi^*_{N,i})$, $\mathbb{P}^*_{Poi}$ is the counterpart of $\mathbb{P}_{Poi}$ conditional on the series of realized samples. By the argument for Theorem 2.2, we have

$$\hat{F}^*_{N,Poi}(z) = \Phi(z) + \frac{\mu^{(3)*}_{N,Poi}}{6(V^*_{Poi})^{3/2}}(1-z^2)\phi(z) + o(n_0^{-1/2}),$$

where $\mu^{(3)*}_{N,Poi} = \sum_{i=1}^{N} E_*\{(I^*_{N,i}(\pi^*_{N,i})^{-1}-1)^3\}(y^*_{N,i})^3$.

It remains to show that

$$\begin{aligned} n_0^2N^{-3}(\mu^{(3)*}_{N,Poi} - \hat{\mu}^{(3)}_{N,Poi}) &\to 0, \qquad (2.40)\\ n_0N^{-2}(V^*_{N,Poi} - \hat{V}_{N,Poi}) &\to 0, \qquad (2.41)\end{aligned}$$

in probability conditional on the series of realized samples. Since the proof for the two results is similar, we prove (2.40). Based on the proposed bootstrap method for Poisson sampling, $N_i^* \sim \text{Bin}(N, \rho_i)$, and recall that $\rho_i = \pi_{N,i}^{-1}(\sum_{j=1}^{n}\pi_{N,j}^{-1})^{-1}$.

The result (2.32) is not sharp enough, and we consider $E\left(1 - N^{-1}\sum_{i=1}^{n}\pi_{N,i}^{-1} \mid \mathcal{F}_N\right) = 0$ and $\text{var}\left(1 - N^{-1}\sum_{i=1}^{n}\pi_{N,i}^{-1} \mid \mathcal{F}_N\right) \le N^{-2}\sum_{i=1}^{N}\pi_{N,i}^{-1} = O(n_0^{-1})$, where the last inequality holds by (C1). Thus, we have

$$1 - N^{-1}\sum_{i=1}^{n}\pi_{N,i}^{-1} = O_p(n_0^{-1/2}). \tag{2.42}$$

Consider

$$\begin{aligned} E_*\{N^{-1}(N_i^* - \pi_{N,i}^{-1})\} &= \pi_{N,i}^{-1}\left\{\left(\sum_{i=1}^{n}\pi_{N,i}^{-1}\right)^{-1} - N^{-1}\right\}\\ &= \pi_{N,i}^{-1}\left(\sum_{i=1}^{n}\pi_{N,i}^{-1}\right)^{-1}\left\{1 - N^{-1}\left(\sum_{i=1}^{n}\pi_{N,i}^{-1}\right)\right\}\\ &= O_p(n_0^{-3/2}) \qquad (2.43)\end{aligned}$$

almost surely based on (2.32) and (2.42). Next, we consider the bootstrap variance of $N^{-1}(N_i^* - \pi_{N,i}^{-1})$. That is,

$$\begin{aligned}\mathrm{var}_*(N^{-1}(N_i^* - \pi_{N,i}^{-1})) &\leq N^{-2}N\pi_{N,i}^{-1}\left(\sum_{i=1}^n \pi_{N,i}^{-1}\right)^{-1} \\ &= O_p(n_0^{-1-1/\alpha}). \end{aligned} \tag{2.44}$$

By (2.43) and (2.44), we have

$$N^{-1}(N_i^* - \pi_{N,i}^{-1}) = O_p(n_0^{-1/2-1/(2\alpha)}). \tag{2.45}$$

By (2.45), we have

$$n_0^{\xi}N^{-1}(N_i^* - \pi_{N,i}^{-1}) = o_p(1) \tag{2.46}$$

if $\xi < 1/2 + 1/(2\alpha)$. Consider

$$\begin{aligned} & n_0^2 N^{-3}(\mu_{N,Poi}^{(3)*} - \hat{\mu}_{N,Poi}^{(3)}) \\ = \;& n_0^2 N^{-2}\sum_{i=1}^n (N^{-1}N_i^* - N^{-1}\pi_{N,i}^{-1})y_{N,i}^3(1-\pi_{N,i})\{(1-\pi_{N,i})^2\pi_{N,i}^{-2} - 1\} \\ = \;& o_p(1)n_0^{-1}\sum_{i=1}^n n_0^{1-\xi}y_{N,i}^3 \\ = \;& o_p(1) \end{aligned} \tag{2.47}$$

for *some* $\xi < 1/2 + 1/(2\alpha)$, and the last equality holds by (C3), (2.37), (2.46) and the fact that $n = \sum_{i=1}^N I_{N,i}$ under Poisson sampling. Thus, we have proved (2.40) by (2.47), and the proof of (2.41) is similar. This concludes the proof of Theorem 2.3 □

Lemma 2.10. *Let i, j, k, l be pairwise distinct positive integers, which are no larger than N. Suppose that (C4) holds. Under SRS, we have*

$$E[(I_{N,i}\pi_{N,i}^{-1} - 1)^4 \mid \mathcal{F}_N] = O(n^{-3}N^3), \tag{2.48}$$

$$E[(I_{N,i}\pi_{N,i}^{-1} - 1)^3(I_{N,j}\pi_{N,j}^{-1} - 1) \mid \mathcal{F}_N] = O(N^2n^{-2}), \tag{2.49}$$

$$E[(I_{N,i}\pi_{N,i}^{-1} - 1)^2(I_{N,j}\pi_{N,j}^{-1} - 1)^2 \mid \mathcal{F}_N] = O(N^2n^{-2}), \tag{2.50}$$

$$E[(I_{N,i}\pi_{N,i}^{-1} - 1)(I_{N,j}\pi_{N,j}^{-1} - 1)(I_{N,k}\pi_{N,k}^{-1} - 1)^2 \mid \mathcal{F}_N] = O(Nn^{-2}), \tag{2.51}$$

$$E[(I_{N,i}\pi_{N,i}^{-1} - 1)(I_{N,j}\pi_{N,j}^{-1} - 1)(I_{N,k}\pi_{N,k}^{-1} - 1)(I_{N,l}\pi_{N,l}^{-1} - 1) \mid \mathcal{F}_N] = O(n^{-2}). \tag{2.52}$$

Proof of Lemma 2.10. For proving the first result, we consider

$$\begin{aligned} E[(I_{N,i}\pi_{N,i}^{-1}-1)^4 \mid \mathcal{F}_N] &= \pi_{N,i}(\pi_{N,i}^{-1}-1)^4 + (1-\pi_{N,i}) \\ &= (1-\pi_{N,i})[(1-\pi_{N,i})^3\pi_{N,i}^{-3}+1] \\ &\le N^3n^{-3}, \end{aligned}$$

where the last inequality holds by the fact that $(1-x)^3+x^3 \le 1$ for $x \in [0,1]$. Thus, we have proved (2.48).

Denote $\#A$ to be the number of elements that equal to 1 in set A. Under SRS, we have

$$\begin{aligned} P_N(\#\{I_{N,i}, I_{N,j}\} = 2) &= \frac{n(n-1)}{N(N-1)}, \\ P_N(\#\{I_{N,i}, I_{N,j}\} = 1) &= \frac{n(N-n)}{N(N-1)}, \\ P_N(\#\{I_{N,i}, I_{N,j}\} = 0) &= \frac{(N-n)(N-n-1)}{N(N-1)}, \end{aligned}$$

Under SRS, we have $\pi_{N,i}^{-1}-1 = (N-n)n^{-1}$ for $i = 1, \ldots, N$. Consider

$$\begin{aligned} & E[(I_{N,i}\pi_{N,i}^{-1}-1)^3(I_{N,j}\pi_{N,j}^{-1}-1) \mid \mathcal{F}_N] \\ = & \frac{(N-n)^4}{n^4}\frac{n(n-1)}{N(N-1)} - \frac{(N-n)^3}{n^3}\frac{n(N-n)}{N(N-1)} - \frac{N-n}{n}\frac{n(N-n)}{N(N-1)} + \frac{(N-n)(N-n-1)}{N(N-1)} \\ = & -\frac{(N-n)^4}{n^3N(N-1)} + O(1) \\ = & O(N^2n^{-2}), \end{aligned} \tag{2.53}$$

where the last inequality holds by the facts that $(N-n)^4[n^3N(N-1)]^{-1} = O(N^2n^{-3})$ and $N^2n^{-2} = O(1)$ if $n = O(N)$. Thus, we have proved (2.49) by (2.53).

Consider

$$\begin{aligned} & E[(I_{N,i}\pi_{N,i}^{-1}-1)^2(I_{N,j}\pi_{N,j}^{-1}-1)^2 \mid \mathcal{F}_N] \\ = & \frac{(N-n)^4}{n^4}\frac{n(n-1)}{N(N-1)} + 2\frac{(N-n)^2}{n^2}\frac{n(N-n)}{N(N-1)} + \frac{(N-n)(N-n-1)}{N(N-1)} \\ = & O(N^2n^{-2}) + O(Nn^{-1}) + O(1) \\ = & O(N^2n^{-2}), \end{aligned}$$

which proves (2.50). Similar with the case for two terms, we have the following results under SRS. That is,

$$\begin{aligned}
P_N(\#\{I_{N,i}, I_{N,j}, I_{N,k}\} = 3) &= \frac{n(n-1)(n-2)}{N(N-1)(N-2)}, \\
P_N(\#\{I_{N,i}, I_{N,j}, I_{N,k}\} = 2) &= \frac{n(n-1)(N-n)}{N(N-1)(N-2)}, \\
P_N(\#\{I_{N,i}, I_{N,j}, I_{N,k}\} = 1) &= \frac{n(N-n)(N-n-1)}{N(N-1)(N-2)}, \\
P_N(\#\{I_{N,i}, I_{N,j}, I_{N,k}\} = 0) &= \frac{(N-n)(N-n-1)(N-n-2)}{N(N-1)(N-2)}.
\end{aligned}$$

Consider

$$\begin{aligned}
& E[(I_{N,i}\pi_{N,i}^{-1} - 1)(I_{N,j}\pi_{N,j}^{-1} - 1)(I_{N,k}\pi_{N,k}^{-1} - 1)^2 \mid \mathcal{F}_N] \\
= \; & \frac{(N-n)^4}{n^4}\frac{n(n-1)(n-2)}{N(N-1)(N-2)} + \frac{(N-n)^2}{n^2}\frac{n(n-1)(N-n)}{N(N-1)(N-2)} - \\
& 2\frac{(N-n)^3}{n^3}\frac{n(n-1)(N-n)}{N(N-1)(N-2)} - 2\frac{N-n}{n}\frac{n(N-n)(N-n-1)}{N(N-1)(N-2)} + \\
& \frac{(N-n)^2}{n^2}\frac{n(N-n)(N-n-1)}{N(N-1)(N-2)} + \frac{(N-n)(N-n-1)(N-n-2)}{N(N-1)(N-2)} \\
= \; & \frac{(n-1)(n-2)(N-n)^4}{n^3N(N-1)(N-2)} + \frac{(N-n)^3}{nN(N-1)} - 2\frac{(n-1)(N-n)^4}{n^2N(N-1)(N-2)} + \\
& \frac{(N-n)(N-n-1)(N-n-2) - 2(N-n)^2(N-n-1)}{N(N-1)(N-2)}.
\end{aligned} \tag{2.54}$$

After some algebra, the first three terms of (2.54) is

$$\begin{aligned}
& \frac{(n-1)(n-2)(N-n)^4}{n^3N(N-1)(N-2)} + \frac{(N-n)^3}{nN(N-1)} - 2\frac{(n-1)(N-n)^4}{n^2N(N-1)(N-2)} \\
= \; & \frac{(N-n)^3}{N(N-1)(N-2)} + O(Nn^{-2}).
\end{aligned} \tag{2.55}$$

Together with (2.54) and (2.55), we have

$$\begin{aligned}
& E[(I_{N,i}\pi_{N,i}^{-1} - 1)(I_{N,j}\pi_{N,j}^{-1} - 1)(I_{N,k}\pi_{N,k}^{-1} - 1)^2 \mid \mathcal{F}_N] \\
= \; & O(Nn^{-2}) - \frac{(N-n)(N-n-2)}{N(N-1)(N-2)} \\
= \; & O(Nn^{-2}),
\end{aligned} \tag{2.56}$$

where the last equality holds by (C4). Thus, we have shown (2.51) by (2.56).

Based on a similar calculation, we can show the following results for the four terms under SRS. That is,

$$
\begin{aligned}
P_N(\#\{I_{N,i}, I_{N,j}, I_{N,k}, I_{N,l}\} = 4) &= \frac{n(n-1)(n-2)(n-3)}{N(N-1)(N-2)(N-3)}, \\
P_N(\#\{I_{N,i}, I_{N,j}, I_{N,k}, I_{N,l}\} = 3) &= \frac{n(n-1)(n-2)(N-n)}{N(N-1)(N-2)(N-3)}, \\
P_N(\#\{I_{N,i}, I_{N,j}, I_{N,k}, I_{N,l}\} = 2) &= \frac{n(n-1)(N-n)(N-n-1)}{N(N-1)(N-2)(N-3)}, \\
P_N(\#\{I_{N,i}, I_{N,j}, I_{N,k}, I_{N,l}\} = 1) &= \frac{n(N-n)(N-n-1)(N-n-2)}{N(N-1)(N-2)(N-3)}, \\
P_N(\#\{I_{N,i}, I_{N,j}, I_{N,k}, I_{N,l}\} = 0) &= \frac{(N-n)(N-n-1)(N-n-2)(N-n-3)}{N(N-1)(N-2)(N-3)}.
\end{aligned}
$$

Now, consider

$$
\begin{aligned}
& E[(I_{N,i}\pi_{N,i}^{-1} - 1)(I_{N,j}\pi_{N,j}^{-1} - 1)(I_{N,k}\pi_{N,k}^{-1} - 1)(I_{N,l}\pi_{N,l}^{-1} - 1) \mid \mathcal{F}_N] \\
= \; & \frac{(N-n)^4}{n^4}\frac{n(n-1)(n-2)(n-3)}{N(N-1)(N-2)(N-3)} - 4\frac{(N-n)^3}{n^3}\frac{n(n-1)(n-2)(N-n)}{N(N-1)(N-2)(N-3)} + \\
& 6\frac{(N-n)^2}{n^2}\frac{n(n-1)(N-n)(N-n-1)}{N(N-1)(N-2)(N-3)} - 4\frac{(N-n)}{n}\frac{n(N-n)(N-n-1)(N-n-2)}{N(N-1)(N-2)(N-3)} + \\
& \frac{(N-n)(N-n-1)(N-n-2)(N-n-3)}{N(N-1)(N-2)(N-3)}.
\end{aligned} \tag{2.57}
$$

Consider

$$
\begin{aligned}
& \frac{(n-1)(n-2)(n-3)}{n^3}(N-n)^4 - \frac{(n-1)(n-2)}{n^2}(N-n)^4 + 6\frac{(n-1)}{n}(N-n)^3(N-n-1) - \\
& 4(N-n)^2(N-n-1)(N-n-2) + (N-n)(N-n-1)(N-n-2)(N-n-3) \\
= \; & \frac{3(N-n)^4}{n^2} - \frac{6(N-n)^4}{n^3} + \frac{6(N-n)^3}{n} + 3(N-n)^2 - 6(N-n) \\
= \; & O(N^4 n^{-2}),
\end{aligned} \tag{2.58}
$$

where the last equality is valid by (C4). Together with (2.57) and (2.58), we have

$$
\begin{aligned}
& E[(I_{N,i}\pi_{N,i}^{-1} - 1)(I_{N,j}\pi_{N,j}^{-1} - 1)(I_{N,k}\pi_{N,k}^{-1} - 1)(I_{N,l}\pi_{N,l}^{-1} - 1) \mid \mathcal{F}_N] \\
= \; & O(N^4 n^{-2})\{N(N-1)(N-2)(N-3)\}^{-1} \\
= \; & O(n^{-2}).
\end{aligned}
$$

Thus, we have proved (2.52). $\square$

Proof of Lemma 2.4. Based on basic algebra and (C2), we can show (2.8) and (2.9), and the proof is omitted here.

Note that $s^2_{N,SRS} = (n-1)^{-1}\sum_{i=1}^n y_i^2 - \frac{n}{n-1}\left(n^{-1}\sum_{i=1}^n y_i\right)^2$, $\sigma^2_{N,SRS} = N^{-1}\sum_{i=1}^N y_i^2 - \left(N^{-1}\sum_{i=1}^N y_i\right)^2$. To show (2.10), it is enough to show that

$$
\begin{aligned}
n^{-1}\sum_{i=1}^n y_i^2 - N^{-1}\sum_{i=1}^N y_i^2 &\rightarrow 0, &(2.59)\\
n^{-1}\sum_{i=1}^n y_i - N^{-1}\sum_{i=1}^N y_i &\rightarrow 0 &(2.60)
\end{aligned}
$$

almost surely. First, we show (2.59), and we have $n^{-1}\sum_{i=1}^n y_i^2 - N^{-1}\sum_{i=1}^N y_i^2 = N^{-1}\sum_{i=1}^N (I_{N,i}\pi_{N,i}^{-1} - 1)y_{N,i}^2$.

Based on (C4) and Lemma 2.1, it is enough to show that $N^{-4}E\left[\left\{\sum_{i=1}^N (I_{N,i}\pi_{N,i}^{-1} - 1)y_{N,i}^2\right\}^4 \mid \mathcal{F}_N\right] = O(n^2)$. By some basic algebra and (C2), we have

$$
\begin{aligned}
\sum_{i\neq j} y_{N,i}^6 y_{N,j}^2 &= O(N^2), &(2.61)\\
\sum_{i\neq j\neq k} y_{N,i}^2 y_{N,j}^2 y_{N,k}^4 &= O(N^3), &(2.62)\\
\sum_{i\neq j\neq k\neq l} y_{N,i}^2 y_{N,j}^2 y_{N,k}^2 y_{N,l}^2 &= O(N^4), &(2.63)
\end{aligned}
$$

where $i \neq j \neq k$ stands for "$\{i,j,k\} \subset \{1,\ldots,N\}$ and they are pairwise distinct", and $i \neq j \neq k \neq l$ for "$\{i,j,k,l\} \subset \{1,\ldots,N\}$, and they are pairwise distinct".

Consider

$$
\begin{aligned}
& N^{-4}E\left[\left\{\sum_{i=1}^N (I_{N,i}\pi_{N,i}^{-1} - 1)y_{N,i}^2\right\}^4 \mid \mathcal{F}_N\right] \\
= \; & N^{-4}\sum_{i=1}^N E\{(I_{N,i}\pi_{N,i}^{-1} - 1)^4 \mid \mathcal{F}_N\} y_{N,i}^8 + N^{-4}\sum_{i\neq j} E\{(I_{N,i}\pi_{N,i}^{-1} - 1)^2 (I_{N,j}\pi_{N,j}^{-1} - 1)^2 \mid \mathcal{F}_N\} y_{N,i}^4 y_{N,j}^4 + \\
& N^{-4}\sum_{i\neq j} E\{(I_{N,i}\pi_{N,i}^{-1} - 1)^3 (I_{N,j}\pi_{N,j}^{-1} - 1) \mid \mathcal{F}_N\} y_{N,i}^6 y_{N,j}^2 + \\
& N^{-4}\sum_{i\neq j\neq k} E\{(I_{N,i}\pi_{N,i}^{-1} - 1)(I_{N,j}\pi_{N,j}^{-1} - 1)(I_{N,k}\pi_{N,k}^{-1} - 1)^2 \mid \mathcal{F}_N\} y_{N,i}^2 y_{N,j}^2 y_{N,k}^4 + \\
& N^{-4}\sum_{i\neq j\neq k\neq l} E\{(I_{N,i}\pi_{N,i}^{-1} - 1)(I_{N,j}\pi_{N,j}^{-1} - 1)(I_{N,k}\pi_{N,k}^{-1} - 1)(I_{N,l}\pi_{N,l}^{-1} - 1) \mid \mathcal{F}_N\} y_{N,i}^2 y_{N,j}^2 y_{N,k}^2 y_{N,l}^2, \\
= \; & O(n^{-2}),
\end{aligned}
$$

where the last equality holds by Lemma 2.10, (2.20) and (2.61) to (2.63). Thus, we have proved (2.59). Similarly, we can prove (2.60).

Note that $\mu^{(3)}_{N,SRS} = N^{-1}\sum_{i=1}^N y^3_{N,i} - 3\bar{Y}_N N^{-1}\sum_{i=1}^N y^2_{N,i} + 2\bar{Y}^3_N$. To show (2.11), it is enough to show

$$n^{-1}\sum_{i=1}^n y_i^3 - N^{-1}\sum_{i=1}^N y^3_{N,i} \quad\rightarrow\quad 0, \tag{2.64}$$

$$n^{-1}\sum_{i=1}^n y_i^2 - N^{-1}\sum_{i=1}^N y^2_{N,i} \quad\rightarrow\quad 0, \tag{2.65}$$

$$n^{-1}\sum_{i=1}^n y_i - N^{-1}\sum_{i=1}^N y_{N,i} \quad\rightarrow\quad 0, \tag{2.66}$$

in probability. We prove (2.64), and the procedure for proving (2.65) and (2.66) is similar. Note that $n^{-1}\sum_{i=1}^n y_i^3 - N^{-1}\sum_{i=1}^N y^3_{N,i} = N^{-1}\sum_{i=1}^N (I_{N,i}\pi^{-1}_{N,i} - 1)y^3_{N,i}$. Consider

$$E\left(n^{-1}\sum_{i=1}^n y_i^3 - N^{-1}\sum_{i=1}^N y^3_{N,i} \mid \mathcal{F}_N\right) \quad=\quad 0 \tag{2.67}$$

$$\begin{aligned} \mathrm{var}\left(n^{-1}\sum_{i=1}^n y_i^3 - N^{-1}\sum_{i=1}^N y^3_{N,i} \mid \mathcal{F}_N\right) &= n^{-1}(1 - nN^{-1})\sigma^2_{N,3} \\ &= O(n^{-1}), \end{aligned} \tag{2.68}$$

where (2.67) holds by the sampling design, $\sigma^2_{N,3}$ is the finite population variance of $\{y^3_{N,1}, \ldots, y^3_{N,N}\}$, and the second equality of (2.68) holds by (C2). Together with (2.67) and (2.68), we have proved (2.65). □

Proof of Theorem 2.6. First, we show that

$$N^{-1}\sum_{i=1}^N (y^*_{N,i})^8 = O_p(1). \tag{2.69}$$

Consider

$$N^{-1}E_*\left\{\sum_{i=1}^N (y^*_{N,i})^8\right\} \quad=\quad n^{-1}\sum_{i=1}^n y^8_{N,i} \tag{2.70}$$

$$\begin{aligned} E\left(n^{-1}\sum_{i=1}^n y^8_{N,i} \mid \mathcal{F}_N\right) &= N^{-1}\sum_{i=1}^N y^8_{N,i} \\ &= O(1). \end{aligned} \tag{2.71}$$

Together with (2.70) and (2.71), we have proved (2.69) using the Markov inequality.

By (C5), there exist a strongly non-latticed distribution $G(x)$ such that

$$G_N(t) \to G(t) \tag{2.72}$$

as $N \to \infty$, where $t \in C(G)$, and $C(G)$ is the set of continuous points of $G(x)$. Next, we show that

$$G_N^*(t) \to G(t) \tag{2.73}$$

in probability for $t \in C(G)$. Notice that

$$G_N^*(t) = N^{-1} \sum_{i=1}^{N} \delta_{(-\infty,t)}(y_{N,i}^*) = N^{-1} \sum_{i=1}^{n} N_i^* \delta_{(-\infty,t)}(y_{N,i}),$$

and recall that $\delta_A(x) = 1$ if $x \in A$ and 0 otherwise for a set A. Thus, we have

$$E_*(G_N^*(t)) = n^{-1} \textstyle\sum_{i=1}^{n} \delta_{(-\infty,t)}(y_{N,i}), \tag{2.74}$$

$$\mathrm{var}_*(G_N^*(t)) \leq N^{-2} \textstyle\sum_{i=1}^{n} N n^{-1} \delta_{(-\infty,t)}(y_{N,i}) = O(N^{-1}), \tag{2.75}$$

$$E\left\{ n^{-1} \textstyle\sum_{i=1}^{n} \delta_{(-\infty,t)}(y_{N,i}) \mid \mathcal{F}_N \right\} = G_N(t), \tag{2.76}$$

$$\mathrm{var}\left\{ n^{-1} \textstyle\sum_{i=1}^{n} \delta_{(-\infty,t)}(y_{N,i}) \mid \mathcal{F}_N \right\} = o(1), \tag{2.77}$$

where results (2.75) and (2.77) are obtained by the fact that $\delta_{(-\infty,t)}(y_{N,i}) \leq 1$ for all t. Together with (2.74) to (2.77), we have

$$G_N^*(t) - G_N(t) \to 0 \tag{2.78}$$

in probability for all t. By (2.72) and (2.78), we have proved (2.73). Thus, by (C4), (2.69) and (2.73), we have

$$\hat{F}_{N,SRS}^*(z) = \Phi(z) + \frac{(1 - 2nN^{-1})\mu_{N,SRS}^{(3)*}}{6\{n(1 - nN^{-1})\}^{1/2}(\sigma_{N,SRS}^*)^3}(1 - z^2)\phi(z) + o_p(n^{-1/2})$$

almost surely conditional on the sample $\{y_1, \ldots, y_n\}$ obtained by SRS in probability, where $\mu_{SRS}^{(3)*}$ and $(\sigma_{SRS}^*)^2$ are the bootstrap central third moment and variance.

Based on Lemma 2.4, it remains to show that

$$\mu_{N,SRS}^{(3)*} - \hat{\mu}_{N,SRS}^{(3)} \quad \to \quad 0 \tag{2.79}$$

$$(\sigma_{N,SRS}^*)^2 - s_{N,SRS}^2 \quad \to \quad 0 \tag{2.80}$$

in probability.By some algebra, it is equivalent to show the following three results. That is,

$$N^{-1}\sum_{i=1}^{N}(y_{N,i}^{*})^{3} - n^{-1}\sum_{i=1}^{n} y_{N,i}^{3} \rightarrow 0, \tag{2.81}$$

$$N^{-1}\sum_{i=1}^{N}(y_{N,i}^{*})^{2} - n^{-1}\sum_{i=1}^{n} y_{N,i}^{2} \rightarrow 0, \tag{2.82}$$

$$N^{-1}\sum_{i=1}^{N} y_{N,i}^{*} - n^{-1}\sum_{i=1}^{n} y_{N,i} \rightarrow 0, \tag{2.83}$$

in probability.

Consider

$$N^{-1}E_{*}\left\{\sum_{i=1}^{N}(y_{N,i}^{*})^{3}\right\} = n^{-1}\sum_{i=1}^{n} y_{N,i}^{3} \tag{2.84}$$

$$\begin{aligned}\mathrm{var}_{*}\left\{N^{-1}\sum_{i=1}^{N}(y_{N,i}^{*})^{3}\right\} &= N^{-2}\sum_{i=1}^{n} Nn^{-1}y_{N,i}^{6} \\ &= o_{p}(1),\end{aligned} \tag{2.85}$$

where the last equality of (2.85) is derived by the Markov inequality and a similar procedure for (2.71). Thus, we have proved (2.81) by (2.84) and (2.85). Similarly, we can prove (2.82) and (2.83). Therefore, we have shown (2.79) and (2.80), which concludes the proof of Theorem 2.6. □

Proof of Lemma 2.7. For simplicity, denote Z_N to be the random variable on $\mathcal{F}_N$ with the probability measure $P_{N,PPS}$, and $Z_{N,i} \sim Z_N$ to be the random variable associated with the sample for $i = 1, \ldots, n$ under PPS sampling. That is, $P_{N,PPS}(Z_N = p_{N,i}^{-1} y_{N,i}) = p_{N,i}$, and recall that $P_{N,PPS}$ is the probability measure defined on $\mathcal{F}_N$ under PPS sampling. Denote the realized value of $Z_{N,i}$ to be $z_{N,i}$.

Thus, we have

$$\begin{aligned}E(N^{-a}|Z_N|^{a} \mid \mathcal{F}_N) &= N^{-a}\sum_{i=1}^{N} p_{N,i}^{-(a-1)}|y_{N,i}|^{a} \\ &= O(1)N^{-1}\sum_{i=1}^{N}|y_{N,i}|^{a} \\ &= O(1),\end{aligned} \tag{2.86}$$

for $a \le 8$, where the second equality holds by (C6), and the last equality holds by (C2). Notice that $N^{-2}\sigma^2_{N,PPS} = E(N^{-2}Z_N^2 \mid \mathcal{F}_N) - \{E(N^{-1}Z_N \mid \mathcal{F}_N)\}^2$, so we have shown (2.14). Similarly, we can show (2.15).

Denote $X^{(3)}_{N,i} = N^{-2}\{Z^2_{N,i} - E(Z^2_N \mid \mathcal{F}_N)\}$ for $i = 1,\dots,n$, and $D^{(3)}_N$ to be the event that $\{n^{-1}|\sum_{i=1}^n X^{(3)}_{N,i}| > \epsilon\}$, where ϵ is a fixed positive number. By the property of PPS sampling, random variables $\{X^{(3)}_{N,i} : i = 1,\dots,n\}$ are pairwise independent. Consider

$$
\begin{aligned}
& E\left\{\left(n^{-1}\sum_{i=1}^n X^{(3)}_{N,i}\right)^4 \mid \mathcal{F}_N\right\} \\
= \;& n^{-4}\left[\sum_{i=1}^n E\{(X^{(3)}_{N,i})^4 \mid \mathcal{F}_N\} + \sum_{i\neq j; i,j=1,\dots,n} E\{(X^{(3)}_{N,i})^2 \mid \mathcal{F}_N\}E\{(X^{(3)}_{N,j})^2 \mid \mathcal{F}_N\}\right] \\
= \;& O(n^{-2}),
\end{aligned}
$$

where the last equality holds by a similar argument with Lemma 2.1 and (2.86). By (C6) and the Borel-Cantelli Lemma, we have shown

$$
n^{-1}N^{-2}\sum_{i=1}^n Z^2_{N,i} - N^{-2}E(Z^2_N \mid \mathcal{F}_N) \to 0 \tag{2.87}
$$

almost surely. Similarly, we can show

$$
n^{-1}N^{-1}\sum_{i=1}^n Z_{N,i} - N^{-1}E(Z_N \mid \mathcal{F}_N) \to 0 \tag{2.88}
$$

almost surely. Note that $N^{-2}s^2_{N,PPS} = N^{-2}(n-1)^{-1}\sum_{i=1}^n z^2_{N,i} - n(n-1)^{-1}(n^{-1}N^{-1}\sum_{i=1}^n z_{N,i})^2$, and $N^{-2}\sigma_{N,PPS} = E(N^{-2}Z^2_N \mid \mathcal{F}_N) - \{E(N^{-1}Z_N \mid \mathcal{F}_N)\}^2$. By (2.87) and (2.88), we have proved (2.16). In order to show (2.17), it is enough to prove

$$
N^{-a}n^{-1}\sum_{i=1}^n Z^a_{N,i} - E(N^{-a}Z^a_N \mid \mathcal{F}_N) = o_p(1) \tag{2.89}
$$

for $a = 1,2,3$. For $a = 1,2,3$, we have

$$
E\left\{N^{-a}n^{-1}\sum_{i=1}^n Z^a_{N,i} - E(N^{-a}Z^a_N \mid \mathcal{F}_N) \mid \mathcal{F}_N\right\} = 0. \tag{2.90}
$$

Consider

$$
\begin{aligned}
\operatorname{var}\left(N^{-a}n^{-1}\sum_{i=1}^n Z^a_{N,i} \mid \mathcal{F}_N\right) &= n^{-1}N^{-2a}E(Z^{2a}_N \mid \mathcal{F}_N) \\
&= o(1),
\end{aligned} \tag{2.91}
$$

where the first equality holds by the property of PPS sampling, and the second one by (2.86). By (2.90) and (2.91), we have proved (2.89), which validates (2.17) of Lemma 2.7. □

Proof of Theorem 2.8. Denote $T_N = N^{-1}(Z_N - Y_N)$. It can be shown that $E(T_N \mid \mathcal{F}_N) = 0$, and $\mathrm{var}(T_N \mid \mathcal{F}_N) = N^{-2}\sigma^2_{N,PPS}$, where $\sigma^2_{N,PPS}$ is the variance of Z_N with respect to $P_{N,PPS}$. Note that $s^2_{N,PPS}$ is an estimator of $\sigma_{N,PPS}$. Therefore, we have $V_{N,PPS}^{-1/2}(\hat{Y}_{N,HH} - Y) = (nN^{-2}\sigma^2_{N,PPS})^{-1/2}\sum_{i=1}^n T_{N,i}$, where $V_{N,PPS} = n^{-1}\sigma^2_{N,PPS}$, $T_{N,i} = N^{-1}(Z_{N,i} - Y_N)$ for $i = 1, \ldots, n$.

Notice that

$$\begin{aligned} |E(T_N^3 \mid \mathcal{F}_N)| &\leq C_{N,7}[E(N^{-3}|Z_N|^3 \mid \mathcal{F}_N) + \{E(N^{-1}|Z_N| \mid \mathcal{F}_N)\}^3] \\ &= O(1), \end{aligned} \tag{2.92}$$

where the inequality holds by the Jensen's inequality (Athreya and Lahiri, 2006, Theorem 3.1.9), and the last equality holds by (2.86). Thus, by (2.92) and Theorem 1 in §16.4 of Feller (2008), we have

$$F_{N,PPS}(z) = \Phi(z) + \frac{\mu^{(3)}_{N,PPS}}{6\sqrt{n}\sigma^3_{N,PPS}}(1 - z^2)\phi(z) + o(n^{-1/2}),$$

where $\mu^{(3)}_{N,PPS} = E\{(Z_N - Y_N)^3 \mid \mathcal{F}_N\}$. By Lemma 2.7 and (2.93), we have shown Theorem 2.8. □

Proof of Theorem 2.9. By a similar argument made for Lemma 2.7, we have

$$n^{-1}N^{-1}\sum_{i=1}^n p^{-1}_{N,a,i} \to 1 \tag{2.93}$$

with probability 1. Recall that $C_N^* = \sum_{i=1}^n N^*_{a,i} p_{N,a,i}$, where $N^*_{a,i}$ is the repetitions in the proposed bootstrap method. Next, we show that

$$C_N^* \quad = \quad 1 + o_p(1) \tag{2.94}$$

conditional on the series of realized samples.

Consider

$$E_*(C_N^*) \quad = \quad nN\left(\sum_{i=1}^n p^{-1}_{N,a,i}\right)^{-1} \tag{2.95}$$

$$\mathrm{var}_*(C_N^*) \quad \leq \quad N\left(\sum_{i=1}^n p^{-1}_{N,a,i}\right)^{-1}\sum_{i=1}^n p_{N,a,i} = O(N^{-1})nN\left(\sum_{i=1}^n p^{-1}_{N,a,i}\right)^{-1}, \tag{2.96}$$

where the equality of (2.96) holds by (C6). By (2.93), (2.95) and (2.96), we have shown (2.94).

Similarly, we can show

$$nN^*_{a,i}p_{N,a,i} = 1 + o_p(1) \tag{2.97}$$

for $i = 1, \ldots, n$ conditional on the series of realized samples.

Conditional on the realized sample, denote $F^*_{N,HH}(z)$ to be the distribution of $(V^*_{N,PPS})^{-1/2}(\hat{Y}^*_{N,HH} - Y^*_N)$, where $\hat{Y}^*_{N,HH} = n^{-1}\sum_{i=1}^n C^*_N(p^*_{N,b,i})^{-1}y^*_{N,b,i}$, $Y^*_N = \sum_{i=1}^N y^*_{N,i}$, $V^*_{PPS} = n^{-1}(\sigma^*_{N,PPS})^2$, $(\sigma^*_{N,PPS})^2$ is the finite population variance of Z^*_N, and $P^*_N\{Z^*_N = C^*_N(p^*_{N,i})^{-1}y^*_{N,i}\} = (C^*_N)^{-1}p^*_{N,i}$ for $i = 1, \ldots, N$. Recall that $\{p^*_{N,i} : i = 1, \ldots, N\}$ consists of $N^*_{a,i}$ copies of $p_{N,a,i}$ for $i = 1, \ldots, n$.

Consider

$$\begin{aligned}
E\left(n^{-1}\sum_{i=1}^n N^{-3}Z^3_{N,i} \mid \mathcal{F}_N\right) &= E(N^{-3}Z^3_N \mid \mathcal{F}_N) \\
&= O(1) &(2.98)\\
\mathrm{var}\left(n^{-1}\sum_{i=1}^n N^{-3}Z^3_{N,i} \mid \mathcal{F}_N\right) &= n^{-1}\mathrm{var}(N^{-3}Z^3_N \mid \mathcal{F}_N) \\
&\leq n^{-1}E(N^{-6}Z^6_N \mid \mathcal{F}_N) \\
&= O(n^{-1}), &(2.99)
\end{aligned}$$

where the results of (2.98) and (2.99) are based on (2.86).

Recall that $E_{**}(\cdot)$ is the expectation with respect to the sampling design conditional on the bootstrap sample. Consider

$$\begin{aligned}
N^{-3}E_{**}\{(Z^*_N)^3\} &= N^{-3}\sum_{i=1}^N (C^*_N)^{-1}p^*_{N,i}\{C^*_N(p^*_{N,i})^{-1}y^*_{N,i}\}^3 \\
&= N^{-3}(C^*_N)^2\sum_{i=1}^N (p^*_{N,i})^{-2}(y^*_{N,i})^3 \\
&= N^{-3}(C^*_N)^2\sum_{i=1}^n N^*_{a,i}p^{-2}_{N,a,i}y^3_{N,a,i} \\
&= \{1 + o_p(1)\}N^{-3}n^{-1}\sum_{i=1}^n p^{-3}_{N,a,i}y^3_{N,a,i} \\
&= O_p(1), &(2.100)
\end{aligned}$$

where the fourth equality holds by (2.97), and last equality holds by Lemma 2.7, (C6), (2.98) and (2.99). Similarly, we can also show

$$N^1 E_{**}(Z_N^*) \quad = \quad \bar{Y}_N^* = O_p(1), \tag{2.101}$$

where $\bar{Y}_N^* = N^{-1}Y_N^*$.

Denote $T_N^* = N^{-1}(Z_N^* - Y_N^*)$. Consider

$$\begin{aligned} N^{-3}E_{**}\{(Z_N^* - Y_N^*)^3\} &\leq C_{N,7}N^{-3}E_{**}\{(Z_N^*)^3\} - (\bar{Y}_N^*)^3 \\ &= O_p(1) \end{aligned} \tag{2.102}$$

by (2.100) and (2.101), where $C_{N,7}$ is a constant with respect to N. Thus, by (2.102) and Theorem 1 in §16.4 of Feller (2008), we have

$$F_{N,PPS}^*(z) = \Phi(z) + \frac{\mu_{N,PPS}^{(3)*}}{6\sqrt{n}(\sigma_{N,PPS}^*)^3}(1 - z^2)\phi(z) + o(n^{-1/2}) \tag{2.103}$$

in probability, where $\mu_{N,PPS}^{(3)*} = E_{**}\{(Z_N^* - Y_N^*)^3\}$.

Next, we show

$$N^{-1}\sum_{i=1}^{N}(y_{N,i}^*)^8 = O_p(1). \tag{2.104}$$

Consider

$$\begin{aligned} E_*\left\{N^{-1}\sum_{i=1}^{N}(y_{N,i}^*)^8\right\} &= \sum_{i=1}^{n} p_{N,a,i}^{-1}\left(\sum_{i=1}^{n} p_{N,a,i}^{-1}\right)^{-1} y_{N,a,i}^8 \\ &\leq C_4^{-1}C_5^8 n^{-1}\sum_{i=1}^{n} N^{-8}Z_{N,i}^8 \end{aligned} \tag{2.105}$$

almost surely by (2.93), where the last equality equality holds by (C6), and recall that $Z_{N,i}$ is the random variable associated with $y_{N,a,i}$ under PPS sampling. By (2.93) and Markov's inequality, we have

$$n^{-1}\sum_{i=1}^{n} N^{-8}Z_{N,i}^8 = O_p(1). \tag{2.106}$$

By (2.105) and (2.106), we have proved (2.104) by the procedure we used for (2.38). Based on a similar argument made for (2.100), we have

$$N^{-3}(\mu_{N,PPS}^{(3)*} - \hat{\mu}_{N,PPS}^{(3)}) \quad = \quad o_p(1), \tag{2.107}$$

$$N^{-2}\{(\sigma_{N,PPS}^*)^2 - s_{N,PPS}^2\} \quad = \quad o_p(1). \tag{2.108}$$

Together with Lemma 2.7, (2.103) to (2.108) and the fact that $N^{-2}\sigma^2_{N,PPS} = O(1)$ derived based on the procedure of (2.92), we have proved Theorem 2.9. □

2.8.2 Estimates for the two-stage sampling design

For the two-stage sampling designs in the second simulation study, Poisson sampling and PPS sampling are used in the first stage, and an SRS design is independently conducted within each selected cluster in the second stage. For the sampling designs in the first stage, denote $\pi_i = n_1 N_i N^{-1}$ to be the first-order inclusion probability for Poisson sampling, and $p_i = N_i N^{-1}$ to be the selection probability for PPS sampling. In this section, we comply to the notation convention in §1.2.8 of Fuller (2009) to discuss the variance estimation under the two-stage sampling designs.

For the two-stage sampling design, where Poisson sampling is applied in the first stage, the design-based estimator of $\bar{Y}$ is

$$\tilde{Y} = N^{-1}\sum_{i\in A}{\pi_i}^{-1}\hat{Y}_{i,\cdot} = n_1^{-1}\sum_{i\in A}N_i^{-1}\hat{Y}_{i,\cdot},$$

where A is the index of the selected clusters in the first stage, $\hat{Y}_{i,\cdot} = N_i n_2^{-1}\sum_{j\in B_i} y_{i,j}$ is an design-unbiased estimate of the cluster total $Y_{i,\cdot} = \sum_{j=1}^{N_i} y_{i,j}$ under the SRS design, and B_i is the index set of the sample within the selected cluster indexed by i. It can be shown that the same form holds when PPS sampling is used in the first stage.

First, we discuss the variance estimator of $\tilde{Y}$ for the two-stage sampling design where Poisson sampling is used in the first stage. As shown in §1.2.8 by Fuller (2009), the variance of $\tilde{Y}$ can be decomposed into two parts. That is,

$$\mathrm{var}(\tilde{Y} \mid U_N) = V_1 + V_2, \tag{2.109}$$

where $V_1 = E[\mathrm{var}\{\tilde{Y} \mid (A, U_N)\} \mid U_N]$ and $V_2 = \mathrm{var}[E\{\tilde{Y} \mid (A, U_N)\} \mid U_N]$.

Consider

$$\mathrm{var}\{\tilde{Y} \mid (A, U_N)\} \quad = \quad N^{-2}\sum_{i\in A}\pi_i^{-2}\mathrm{var}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}, \tag{2.110}$$

where the equality holds since the SRS design is independently conducted within each selected cluster, $\mathrm{var}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\} = N_i(N_i - n_2)n_2^{-1}S_i^2$, $S_i^2 = (N_i - 1)^{-1}\sum_{j=1}^{N_i}(y_{i,j} - \bar{Y}_{i,\cdot})^2$ is the finite population variance within the i-th cluster, and $\bar{Y}_{i,\cdot} = N_i^{-1}Y_{i,\cdot}$ is the finite population mean of the i-th cluster. Since the sample variance $s_i^2 = (n_2 - 1)^{-1}\sum_{j\in B_i}(y_{i,j} - \tilde{Y}_{i,\cdot})^2$ is an unbiased estimator of S_i^2, where $\tilde{Y}_{i,\cdot} = N_i^{-1}\hat{Y}_{i,\cdot}$ is the estimated cluster mean, the first term of (2.109) can be estimated by

$$\hat{V}_1 = N^{-2}\sum_{i\in A}\pi_i^{-2}\hat{V}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}, \tag{2.111}$$

where $\hat{V}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\} = N_i(N_i - n_2)n_2^{-1}s_i^2$.

For the second term of (2.109), consider $E\{\tilde{Y} \mid (A, U_N)\} = N^{-1}\sum_{i\in A}\pi_i^{-1}Y_{i,\cdot}$. Since Poisson sampling is used in the first stage, we have

$$\mathrm{var}[E\{\tilde{Y} \mid (A, U_N)\} \mid U_N] \;=\; N^{-2}\sum_{i=1}^{H}\pi_i^{-1}(1 - \pi_i)Y_{i,\cdot}^2, \tag{2.112}$$

which can be estimated by $N^{-2}\sum_{i\in A}\pi_i^{-2}(1 - \pi_i)Y_{i,\cdot}^2$. Notice that

$$\begin{aligned} E\{\hat{Y}_{i,\cdot}^2 \mid (A, U_N)\} &= [E\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}]^2 + \mathrm{var}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\} \\ &= Y_{i,\cdot}^2 + \mathrm{var}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}. \end{aligned} \tag{2.113}$$

By (2.112) and (2.113) and the fact that s_i^2 is an unbiased estimator of S_i^2, the second term of (2.109) can be estimated by

$$\hat{V}_2 = N^{-2}\sum_{i\in A}\pi_i^{-2}(1 - \pi_i)[\hat{Y}_{i,\cdot}^2 - \hat{V}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}]. \tag{2.114}$$

By (2.111) and (2.114), the variance of $\tilde{Y}$ can be estimated by $\tilde{V} = N^{-2}[\sum_{i\in A}\pi_i^{-2}(1 - \pi_i)\hat{Y}_{i,\cdot}^2 + \sum_{i\in A}\pi_i^{-1}\hat{V}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}]$, when Poisson sampling is used in the first stage.

Next, we use variance decomposition (2.109) to derive the variance estimator of $\tilde{Y}$ under the two-stage sampling design where PPS sampling is applied in the first stage. The result shown in (2.110) holds, and we can still use (2.111) to approximate V_1.

Consider

$$\mathrm{var}[E\{\tilde{Y} \mid (A, U_N)\} \mid U_N] \;=\; N^{-2}n_1^{-1}(n_1 - 1)^{-1}\left(\sum_{i\in A}Z_{i,\cdot}^2 - n_1\bar{Z}^2\right), \tag{2.115}$$

where the equality holds by the property of PPS sampling, $Z_{i,\cdot} = Y_{i,\cdot} p_i^{-1}$ and $\bar{Z} = n_1^{-1} \sum_{i \in A} Z_{i,\cdot}$. Based on (2.113), we can estimate $Z_{i,\cdot}^2$ by

$$p_i^{-2}[\hat{Y}_{i,\cdot}^2 - \hat{V}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}].$$

Consider $E\left\{\tilde{Z}^2 \mid (A, U_N)\right\} = \bar{Z}^2 + \text{var}\{\tilde{Z} \mid (A, U_N)\} = \bar{Z}^2 + n_1^{-2} \sum_{i \in A} p_i^{-2} \text{var}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}$, where $\tilde{Z} = n_1^{-1} \sum_{i \in A} \hat{Y}_{i,\cdot} p_i^{-1}$. Thus, we can estimate $\bar{Z}^2$ by

$$\tilde{Z}^2 - n_1^{-2} \sum_{i \in A} p_i^{-2} \hat{V}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}.$$

By (2.111), (2.115) and the two approximations above, we can obtain the variance estimate of $\tilde{Y}$ by

$$\tilde{V} \;=\; N^{-2} n_1^{-1} (n_1 - 1)^{-1} \sum_{i \in A} p_i^{-2} [\hat{Y}_{i,\cdot}^2 + (n_1 - 2) n_1^{-1} \hat{V}\{\hat{Y}_{i,\cdot} \mid (A, U_N)\}] - N^{-2} \tilde{Z}^2$$

for the two-stage sampling design with PPS sampling is used in the first stage.

References

Athreya, K. B. and Lahiri, S. N. (2006). *Measure theory and probability theory.* Springer Science & Business Media, New York.

Bartholdi, J. J. and Platzman, L. K. (1988). Heuristics based on spacefilling curves for combinatorial problems in euclidean space. *Management Science*, 34(3):291–305.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45.

Breidt, F. J. (1996). Markov chain designs for one-per-stratum sampling. *Survey Methodology*, 21(1):63 – 70.

Cochran, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *The Annals of Mathematical Statistics*, 17(2):164–177.

Cressie, N. (2015). *Statistics for spatial data.* John Wiley & Sons, New York, revised edition.

Doukhan, P. (1994). *Mixing. Properties and Examples. Lecture Notes in Statistics.* Springer, New York.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap.* CRC press, Florida.

Grafström, A., Lundström, N. L. P., and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2):514–520.

Grafström, A. and Schelin, L. (2014). How to select representative samples. *Scandinavian Journal of Statistics*, 41(2):277–290.

Grenander, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *The Annals of Mathematical Statistics*, 25(2):252–272.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic press, New York.

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382.

Klenke, A. (2014). *Probability Theory: A Comprehensive Course.* Universitext. Springer-Verlag London Ltd., London, second edition.

Koul, H. L. (1992). M-estimators in linear models with long range dependent errors. *Statistics & probability letters*, 14(2):153–164.

Lahiri, S. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhyā: The Indian Journal of Statistics*, 65(2):356–388.

Lahiri, S. and Mukherjee, K. (2004). Asymptotic distributions of M-estimators in a spatial regression model under some fixed and stochastic spatial sampling designs. *Annals of the Institute of Statistical Mathematics*, 56(2):225–250.

Lahiri, S. N., Kaiser, M. S., Cressie, N., and Hsu, N.-J. (1999). Prediction of spatial cumulative distribution functions using subsampling. *Journal of the American Statistical Association*, 94(445):86–97.

Lahiri, S. N. and Zhu, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *The Annals of Statistics*, 34(4):1774–1813.

Lister, A. J. and Scott, C. T. (2009). Use of space-filling curves to select sample locations in natural resource monitoring studies. *Environmental monitoring and assessment*, 149(1):71–80.

Menezes, R., Garcia-Soidan, P., and Ferreira, C. (2010). Nonparametric spatial prediction under stochastic sampling design. *Journal of Nonparametric Statistics*, 22(3):363–377.

Munholland, P. L. and Borkowski, J. J. (1996). Simple latin square sampling+ 1: A spatial design using quadrats. *Biometrics*, 52(1):125–136.

Nordman, D. J. and Lahiri, S. N. (2004). On optimal spatial subsample size for variance estimation. *Annals of statistics*, 32(5):1981–2027.

Nordman, D. J., Lahiri, S. N., and Fridley, B. L. (2007). Optimal block size for variance estimation by a spatial block bootstrap method. *Sankhyā: The Indian Journal of Statistics*, 69(3):468–493.

Papageorgiou, I. and Karakostas, K. (1998). On optimal sampling designs for autocorrelated finite populations. *Biometrika*, 85(2):482–486.

Politis, D. N., Paparoditis, E., and Romano, J. P. (1998). Large sample inference for irregularly spaced dependent observations based on subsampling. *Sankhyā: The Indian Journal of Statistics, Series A*, 60(2):274–292.

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235.

Stevens, D. L. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, 99(465):262–278.

Yajima, Y. (1991). Asymptotic properties of the lse in a regression model with long-memory stationary errors. *The Annals of Statistics*, 19(1):158–177.

CHAPTER 3. RESAMPLING METHODS FOR ONE-PER-STRATUM SPATIAL SAMPLING DESIGN

Zhonglei Wang and Zhengyuan Zhu

Abstract

In areal sampling, one-per-stratum design is a common approach, which can achieve spatial balance and improve the precision of the resulting estimators. The downside of such design is that it is challenging to have a good design-unbiased variance estimation. In this paper, we propose a general class of stratified sampling design, by which a spatially balanced sample can be generated. The sample is used to get the M-estimator of the coefficients in a spatial linear regression model, and a resampling approach is used to obtain the corresponding variance estimate. Asymptotic properties of the M-estimator and resampling based variance estimator under the proposed design are studied. Simulations are conducted to test the spatial balance of the sample generated by proposed design and the resampling method.

Key Words: Asymptotics; M-estimator; Spatially balanced; Variance estimation.

3.1 Introduction

In environmental studies, populations are often weak dependent in the sense that the correlation between two observations is a function of the their displacement, and it decreases to zero rapidly as their distance increases (Doukhan, 1994). For the weak dependent population, it is desirable to obtain a spatially balanced sample, the one that spreads over the sampling domain well, to make efficient inference (Cochran, 1946; Stevens and Olsen, 2004; Grafström et al., 2012).

Cochran (1946) studied the relative efficiency of systematic and stratified designs in the one-dimensional space. Papageorgiou and Karakostas (1998) discussed the optimal sampling design when the autocorrelation function is integer convex. To generalize stratified designs to two-dimensional space, Munholland and Borkowski (1996) used a simple Latin square to draw a sample from nonoverlapping sampling units, and they demonstrated that estimators by this design is generally more efficient than those by simple random sampling. Breidt (1996) discussed a Markov chain sampling design based on regular grids. Stevens and Olsen (2004) introduced a generalized random tessellation stratified design (GRTS), and a sample is generated by random quadrant-recursive maps. Stevens and Olsen (2004) showed that the sample generated by GRTS is more spatially balanced than that by simple random sampling. Lister and Scott (2009) and Bartholdi and Platzman (1988) used space-filling curves to obtain spatially balanced samples. Grafström et al. (2012) proposed the local pivotal method (LPM), which generates a sample by updating the first-order inclusion probabilities iteratively. Grafström et al. (2012) shown that LPM can be extended to a higher-dimensional space easily, and the sample generated by LPM is more spatially balanced than that by GRTS. However, for most spatially balanced sampling designs, the design-unbiased variance estimation is intractable. Some compromising methods were proposed, such as the local approximation approach (Stevens and Olsen, 2004; Grafström and Schelin, 2014) and the variance estimation under specific designs (Grafström et al., 2012).

Bootstrap (Efron, 1979; Efron and Tibshirani, 1994) is widely used for making inference. Specific for the spatial settings, Lahiri et al. (1999) used a subsampling method to estimate the spatial cumulative distribution function based on the observations on hexagonal grids. Nordman and Lahiri (2004) and Nordman et al. (2007) discussed a block bootstrap method to estimate the variance when observations are regularly spaced. Politis et al. (1998) proposed a subsample approach for observations generated by a homogeneous Poisson process. Lahiri and Zhu (2006) discussed both fixed and stochastic sampling designs, and a block resampling method was used to make inference. Shao (2010) introduced a wild bootstrap method for irregular spaced time series, and argued that this method can be generalized to higher-dimensional space. However, the sample generated by the stochastic sampling designs mentioned above can be regarded as independent and identically distributed with respect to a fixed sampling density.

In this paper, we propose a one-per-stratum sampling design, study the asymptotic properties of the M-estimator under a spatial setting (Grenander, 1954; Koul, 1992; Yajima, 1991) based on the sample generated by this design, and extend the resampling method discussed by Lahiri and Zhu (2006) to make inference. We argue that the proposed design can generate a spatially balanced sample, and an asymptotically unbiased variance estimator can be obtained by the resampling method. We only consider the pure increasing domain asymptotic structure (Cressie, 2015§2.6.3) to avoid the case where sampled locations are close to each other. However, the asymptotic properties for mixed-increasing domain can be derived if certain conditions hold, which is not the topic of this paper. The proposed design is flexible, and the resampling method can be easily implemented in practice.

3.2 One-per-stratum sampling design

Consider a d-dimensional Euclidean space $\mathbb{R}^d$, where $d \geq 2$. Let $R_0 \subset (-1/2, 1/2]^d$ be a Borel set containing the origin as its inner point. The sampling domain, denoted as R_n, is obtained by inflating R_0. That is,

$$R_n = \lambda_n R_0,$$

where n is the sample size and $\lambda_n = O(n^{1/d})$; that is, we only consider the pure increasing domain asymptotic structure. Denote $\mathcal{A}_n = \{A_i : i = 1, \ldots, n\}$ to be the set of pairwise disjoint partition regions such that

$$R_n = \cup_{i=1}^n A_i$$

for $n \geq 1$. Let $\mathcal{S}_n = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ be a sample of size n, where $\boldsymbol{s}_i$ is generated by a sampling density $f_i(\boldsymbol{s})$ on A_i independently for $i = 1, \ldots, n$. We implicitly assume that A_i and $\boldsymbol{s}_i$ are indexed by n to make the notation simpler.

Denote $(\Omega_n, \mathcal{F}_n, P_n)$ to be the probability space corresponding to the sampling procedure, where $\Omega_n = \times_{i=1}^n A_i$ is the product of $A_1, \ldots, A_n$, $\mathcal{F}_n = \times_{i=1}^n \mathcal{G}_i$ is the product σ-algebra of $\mathcal{G}_1, \ldots, \mathcal{G}_n$, P_n is the product probability measure of $P_{n,1}, \ldots, P_{n,n}$, and $(A_i, \mathcal{G}_i, P_{n,i})$ is a probability space with $P_{n,i}$ being the probability measure with respect to $f_i(\boldsymbol{s})$ for $i = 1, \ldots, n$; see Athreya and Lahiri (2006) for details. For any finite set $J \subset \mathbb{N}_+$, let $\mathbb{P}_J$ be the product probability measure on the product measurable space $(\times_{j \in J}\Omega_j, \times_{j \in J}\mathcal{F}_j)$, where $\mathbb{N}_+$ is the set of positive integers. It can be shown that $\{\mathbb{P}_J : J \in \mathbb{N}_+\}$ is a consistent family of finite-dimensional distributions (Klenke, 2014). Thus, by the Kolmogorov's consistency theorem, there exists a probability measure $P_{\boldsymbol{S}}$ on the product space $\Omega_{\boldsymbol{S}} = \times_{n=1}^{\infty}\Omega_n$ equipped with the product σ-algebra, such that $\mathbb{P}_J = P_{\boldsymbol{S}} \circ \xi_J^{-1}$ for all finite positive integer set $J \subset \mathbb{N}_+$, where ξ_J is the canonical projection from $\Omega_{\boldsymbol{S}}$ to the product space $\times_{j \in J}\Omega_j$.

3.3 M-estimator in spatial linear regression models

Consider the following spatial multivariate linear regression model (Yajima, 1991; Lahiri and Zhu, 2006), that is,

$$Y(\boldsymbol{s}) = \boldsymbol{w}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta} + Z(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d, \tag{3.1}$$

where $\boldsymbol{w}(\boldsymbol{s})$ is a known p-dimensional real-valued function, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the coefficient vector, $\{Z(\boldsymbol{s})\}$ is a one-dimensional random field on $\mathbb{R}^d$, and A^{T} is the transpose of a matrix A. We observe $\{Y(\boldsymbol{s}_i) : \boldsymbol{s}_i \in \mathcal{S}_n\}$, and the parameter of interest is $\boldsymbol{\beta}$. Let

$$M_n(\boldsymbol{x}) = \sum_{i=1}^{n} \boldsymbol{w}(\boldsymbol{s}_i)\Psi\{Y(\boldsymbol{s}_i) - \boldsymbol{w}(\boldsymbol{s}_i)^{\mathrm{T}}\boldsymbol{x}\}, \tag{3.2}$$

where $\Psi : \mathbb{R} \to \mathbb{R}$ is a one-dimensional known non-decreasing Borel-measurable function, and $\Psi\{Z(\boldsymbol{s})\}$ is a zero-mean stationary random field on $\mathbb{R}^d$. The M-estimator $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ is obtained as a solution of the following equation

$$M_n(\boldsymbol{x}) = \mathbf{0}, \tag{3.3}$$

where $\mathbf{0}^{\mathrm{T}} = (0, \ldots, 0)$ is a vector of length p.

Before exploring the asymptotic properties of $\hat{\boldsymbol{\beta}}_n$, we introduce the strong mixing condition for the stationary random field $\{\Psi\{Z(\boldsymbol{s})\} : \boldsymbol{s} \in \mathbb{R}^d\}$. For $\boldsymbol{s}^{\mathrm{T}} = (s_1, \ldots, s_d)$, let $\|\boldsymbol{s}\|_1 = \sum_{i=1}^{d} |s_i|$ and $\|\boldsymbol{s}\|_2 = (s_1^2 + \cdots + s_d^2)^{1/2}$. Denote vol.$(A)$ to be the volume of set $A \subset \mathbb{R}^d$, and $|A|$ to be the cardinality of A. For two sets T_1 and T_2 of $\mathbb{R}^d$, let $d(T_1, T_2) = \inf\{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|_2 : \boldsymbol{s}_i \in T_i, i = 1, 2\}$.

Denote $\mathcal{R}(b) = \{\cup_{i=1}^{k} D_i : \sum_{i=1}^{k} \text{vol.}(D_i) \leq b, k \geq 1\}$, where $\{D_i : i = 1, \ldots, k\}$ is a finite set of pairwise disjoint hypercubes in $\mathbb{R}^d$ for $k \in \mathbb{N}_+$. For $a > 0, b > 0$, the strong-mixing coefficient for the random field $\{\Psi\{Z(\boldsymbol{s})\} : \boldsymbol{s} \in \mathbb{R}^d\}$ is defined as

$$\alpha(a; b) = \sup\{\tilde{\alpha}(T_1, T_2); d(T_1, T_2) \geq a, T_1 \in \mathcal{R}(b), T_2 \in \mathcal{R}(b)\},$$

where $\tilde{\alpha}(T_1, T_2) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_Z(T_1), B \in \mathcal{F}_Z(T_2)\}$, $\mathcal{F}_Z(A) = \sigma\langle \Psi\{Z(\boldsymbol{s})\} : \boldsymbol{s} \in A\rangle$ is the sigma-algebra generated by the stationary random field on region $A \subset \mathbb{R}^d$. Assume

$$\alpha(a; b) \leq \alpha_1(a) g_1(b),$$

where $\alpha_1(\cdot)$ is a nonincreasing left continuous function such that $\lim_{a\to\infty} \alpha_1(a) = 0$, and $g_1(\cdot)$ is a nondecreasing function with $\lim_{b\to\infty} g_1(b) = \infty$.

3.4 Asymptotic properties of the M-estimator

We need the following regular conditions on the prototype, partition set and function $\Psi(\cdot)$.

(C1) The prototype R_0 satisfies $\text{vol.}(R_0) > 0$ and $\text{vol.}(R_0^{\epsilon_n}) \to 0$, where $R_0^{\epsilon} = \{x \in R_0 : (x + \epsilon[-1, 1]^d) \cap R_0^{\mathrm{C}} \neq \emptyset\}$, A^{C} is the complement of a set A, and $\epsilon_n \to 0$ as $n \to \infty$.

(C2) There exists $M_A \in (0, \infty)$ such that $\text{vol.}(A_i) \leq M_A$ for $i = 1, \ldots, n$ and $n \in \mathbb{N}_+$.

(C3) There exists $C_1 \in (0, \infty)$ such that $|\{A_i : A_i \cap B \neq \emptyset, i = 1, \ldots, n\}| \leq C_1 \text{vol.}(B)$ for any $B \subset R_n$.

(C4) There exists $M_f \in (0, \infty)$ such that $f_i(\boldsymbol{s}) \leq M_f$ for $\boldsymbol{s} \in \text{supp}(f_i) \subset A_i$, $i = 1, \ldots, n$ and $n \in \mathbb{N}_+$, where $\text{supp}(f) = \{x : f(x) > 0\}$ is the support of a function $f(x)$.

(C5) There exists a sequence of nonsingular matrices $\{\Lambda_n : n \geq 1\}$ such that

$$\Lambda_n^{-1} \left\{ \sum_{i=1}^{n} \int \boldsymbol{w}(\boldsymbol{s}) \boldsymbol{w}(\boldsymbol{s})^{\mathrm{T}} f_i(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} \right\} \Lambda_n^{-1} \to H \text{ as } n \to \infty,$$

and for any $\boldsymbol{h} \in \mathbb{R}^d$,

$$\Lambda_n^{-1} \left\{ \sum_{i=1}^{n} \sum_{j \neq i} \int \boldsymbol{w}(\boldsymbol{y} + \boldsymbol{h}) \boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}} f_i(\boldsymbol{y} + \boldsymbol{h}) f_j(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} \right\} \Lambda_n^{-1} \to Q(\boldsymbol{h}) \text{ as } n \to \infty,$$

where H is a positive definitive matrix, and $Q(\boldsymbol{h})$ is a $p \times p$ matrix-valued function on $\mathbb{R}^d$.

(C6) $\int Q(\boldsymbol{h})\sigma_\Psi(\boldsymbol{h})\mathrm{d}\boldsymbol{h}$ is positive definite, where $\sigma_\Psi(\boldsymbol{h}) = E[\Psi\{Z(\mathbf{0})\}\Psi\{Z(\boldsymbol{h})\}]$.

(C7) $m_{0n} = \sup\{\|\Lambda_n^{-1}\boldsymbol{w}(\boldsymbol{s})\| : \boldsymbol{s} \in R_n\} = o(n^{-3/8})$.

(C8) There exists $\delta \in (0, \infty)$ such that

(C8.1) $E|\Psi\{Z(\mathbf{0})\}|^{2+\delta} < \infty$, $E|\Psi'\{Z(\mathbf{0})\}|^{2+\delta} < \infty$, where $\Psi'(x)$ is the first derivative of $\Psi(x)$. Assume $\chi_0 = E\Psi'\{Z(\mathbf{0})\} \neq 0$.

(C8.2) $\alpha_1(a) = O(a^{-\tau})$, where $\tau > d(2+\delta)/\delta$.

(C8.3) $g_1(b) = o(b^{(\tau-d)/(4d)})$.

(C9) Function $\Psi'(x)$ satisfies a Lipschitz condition of order $\gamma \in (2/3, 1]$, that is,

$$|\Psi'(x_1) - \Psi'(x_2)| \leq C_2|x_1 - x_2|^\gamma,$$

where $x_i \in \mathbb{R}$ for $i = 1, 2$, and C_2 is a fixed constant.

Condition (C1) is a mild restriction on the boundary of the prototype R_0, and R_0 can take commonly used shapes, such as hyperparallelepiped, hyperellipsoid and some nonconvex sets. Conditions (C2) and (C3) regulate the partition regions, and they guarantee that the sample generated by the one-per-stratum design is spatially balanced. By (C1) and (C3), we can show that the number of partition regions on the "boundary" part of R_n is negligible compared with that in its "inner" part, and this result is used to derive the asymptotic properties for the resampling method. Condition (C4) is a mild restriction on the sampling design, and most distributions satisfy this condition. We do not require $\mathrm{supp}(f_i) = A_i$, so a more spatially balanced sample can be generated using a sampling density $f_i(\boldsymbol{s})$ with a more concentrated support. Condition (C5) is the commonly used Grenader condition for the linear regression model (Grenander, 1954) under the proposed design. Condition (C6) guarantees the existence of the variance matrix for the M-estimator $\hat{\boldsymbol{\beta}}_n$. Conditions (C7) regulates the covariate $\boldsymbol{w}(\boldsymbol{s})$, and it is used to show the convergence of relevant statistics. Condition (C8) is needed to show that a central limit theorem holds for the stationary spatial process. Condition (C9) is used for the Taylor's expansion when deriving the asymptotic property of $M_n(\boldsymbol{\beta})$.

First, we establish the asymptotic property $M_n(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is the true coefficient vector in the model (3.1).

Theorem 3.1. *Suppose that (C2) to (C9) hold. For any unit vector $\boldsymbol{a} \in \mathbb{R}^p$,*

$$\boldsymbol{a}^{\mathrm{T}} \Lambda_n^{-1} M_n(\boldsymbol{\beta}) \xrightarrow{d} N(0, \sigma_{\boldsymbol{a}}^2) \quad a.s. \ (P_{\boldsymbol{S}}), \tag{3.4}$$

where "$\xrightarrow{d}$" is short for "converge in conditional distribution given sampled locations $\mathcal{S}_n$", and $\sigma_{\boldsymbol{a}}^2 = \boldsymbol{a}^{\mathrm{T}} H \sigma_\Psi(\mathbf{0}) \boldsymbol{a} + \boldsymbol{a}^{\mathrm{T}} \left\{ \int \sigma_\Psi(\boldsymbol{h}) Q(\boldsymbol{h}) \mathrm{d}\boldsymbol{h} \right\} \boldsymbol{a}$.

By the Cramer-Wold device (Athreya and Lahiri, 2006, Theorem 10.4.5), we have the following corollary.

Corollary 3.2. *Suppose the conditions in Theorem 3.1 hold. Then, we have*

$$\Lambda_n^{-1} M_n(\boldsymbol{\beta}) \xrightarrow{d} N(0, \Sigma_M) \quad a.s. \ (P_{\boldsymbol{S}}), \tag{3.5}$$

where $\Sigma_M = H \sigma_\Psi(\mathbf{0}) + \int \sigma_\Psi(\boldsymbol{h}) Q(\boldsymbol{h}) \mathrm{d}\boldsymbol{h}$.

Corollary 3.2 shows that the distribution of the estimating function $M_n(\boldsymbol{\beta})$ is asymptotically normal almost surely after rescaling by Λ_n, and its variance is determined by the one-per-stratum sampling design and the dependence structure of $\Psi(x)$.

Theorem 3.3. *Suppose that (C2) to (C9) hold. Then,*

$$\Lambda_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} N(0, \chi_0^{-2} \Sigma_\beta) \quad a.s. \ (P_{\boldsymbol{S}}), \tag{3.6}$$

where $\Sigma_\beta = H^{-1} \sigma_\Psi(\mathbf{0}) + H^{-1} \left\{ \int \sigma_\Psi(\boldsymbol{h}) Q(\boldsymbol{h}) \mathrm{d}\boldsymbol{h} \right\} H^{-1}$.

Theorem 3.3 shows that the asymptotic distribution of the M-estimator $\hat{\boldsymbol{\beta}}_n$ is normal given $\mathcal{S}_n = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ almost surely. For the case that there are more than one solutions to (3.3), Lahiri and Zhu (2006) gave a comprehensive consideration, which is also applied under the proposed design. However, the asymptotic variance component Σ_β in Theorem 3.3 is intractable in practice, so we generalize the resampling method discussed by Lahiri and Zhu (2006) to the proposed design in next section.

Denote $g(\boldsymbol{s})$ to be a probability density function on R_0, and $\{\boldsymbol{X}_i : i = 1, 2, \ldots\}$ to be independent and identically distributed random vectors with the density $g(\boldsymbol{s})$. Assume that $\{\boldsymbol{X}_i : i = 1, 2, \ldots\}$ is independent with $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in R_n\}$ for $n \geq 1$. A sample for R_n is obtained by $\boldsymbol{s}_i = \lambda_n \boldsymbol{x}_i$, where $\boldsymbol{x}_i$ is a realization of $\boldsymbol{X}_i$ on R_0. Such sampling design is also considered by Shao (2010) and Menezes et al. (2010). As it is often assumed that the selection probability for each location is the same, we compare the asymptotic efficiency of the M-estimator under the proposed design and that considered by Lahiri and Zhu (2006) using sampling density $g(\boldsymbol{s}) = \{\text{vol.}(R_0)\}^{-1}$ for $\boldsymbol{s} \in R_0$. For the proposed design, the partition regions satisfy (C2) and (C3), and have the same volume. Since a sample of size n is selected from R_n, we have $\text{vol.}(A_i) = \lambda_n^d \text{vol.}(R_0)/n \to \text{vol.}(R_0)/c$ as $n \to \infty$, where $c \in (0, \infty)$ under the pure increasing domain asymptotic structure. Define the sampling density $f_i(\boldsymbol{s}) = n\{\lambda_n^d \text{vol.}(R_0)\}^{-1} \mathbb{1}(\boldsymbol{s} \in A_i)$ for $i = 1, \ldots, n$.

Theorem 3.4. *Assume that a uniform distribution is used to draw a sample under the stochastic design discussed by Lahiri and Zhu (2006), and a uniform distribution is used for each partition region under the proposed design, where the partition regions satisfy (C2) and (C3), and have the same volume. Furthermore, suppose that conditions (C5) to (C9) hold, and there exists a sequence of nonsingular matrices $\{\Lambda_{n,iid}\}$ such that*

$$\Lambda_{n,iid}^{-1} \left\{ \frac{1}{\text{vol.}(R_0)} \int_{R_0} \boldsymbol{w}(\lambda_n \boldsymbol{s}) \boldsymbol{w}(\lambda_n \boldsymbol{s})^{\mathrm{T}} \mathrm{d}\boldsymbol{s} \right\} \Lambda_{n,iid}^{-1} \to H_{iid} \quad \text{as } n \to \infty, \tag{3.7}$$

$$\Lambda_{n,iid}^{-1} \left\{ \frac{1}{\{\text{vol.}(R_0)\}^2} \int_{R_0} \boldsymbol{w}(\lambda_n \boldsymbol{s} + \boldsymbol{h}) \boldsymbol{w}(\lambda_n \boldsymbol{s})^{\mathrm{T}} \mathrm{d}\boldsymbol{s} \right\} \Lambda_{n,iid}^{-1} \to Q_{iid}(\boldsymbol{h}) \quad \text{as } n \to \infty, \tag{3.8}$$

where H_{iid} is a positive definite matrix and $Q_{iid}(\boldsymbol{h})$ is a $p \times p$ matrix-valued function, and $\boldsymbol{w}(\boldsymbol{s}) = 0$ if $\boldsymbol{s} \notin R_n$.

Then, the M-estimator of $\boldsymbol{\beta}$ from this one-per-stratum sampling design is at least as efficient as the independent and identically distributed stochastic sampling design asymptotically in the sense that the asymptotic variance of any element of the scaled $\hat{\boldsymbol{\beta}}_n$ under the one-per-stratum sampling design is no larger than that under the independent and identically distributed sampling design.

Furthermore, if $\int_A \boldsymbol{w}(\lambda_n \boldsymbol{s} + \boldsymbol{h})\boldsymbol{w}(\lambda_n \boldsymbol{s})^{\mathrm{T}} \mathrm{d}\boldsymbol{s}$ *is positive definitive for any* $A \subset \mathcal{B}(\mathbb{R}^d \cap R_n)$ *with a positive Lebesgue measure, the M-estimator from the corresponding one-per-stratum sampling design is more efficient asymptotically.*

Theorem 3.4 demonstrates that the proposed design is more efficient for estimating the parameters in spatial linear regression problem than the stochastic design considered by Lahiri and Zhu (2006) under mild conditions.

3.5 Resampling method

To make inference for the M-estimator $\hat{\boldsymbol{\beta}}_n$ under the proposed design, we generalize the resampling method discussed by Lahiri and Zhu (2006).

Let $\mathcal{K}_n = \{\boldsymbol{k} \in \mathbb{Z}^d : (\boldsymbol{k}b_n + [0,1)^d b_n) \cap R_n \neq \emptyset\} = \mathcal{K}_{1n} \cup \mathcal{K}_{2n}$, where b_n is the block size satisfying certain conditions, $\mathcal{K}_{1n} = \{\boldsymbol{k} \in \mathbb{Z}^d : (\boldsymbol{k}b_n + [0,1)^d b_n) \subset R_n\}$, and $\mathcal{K}_{2n} = \mathcal{K}_n \cap \mathcal{K}_{1n}^{\mathrm{C}}$. The sampling region R_n can be partitioned by $\{R_n(\boldsymbol{k}) : \boldsymbol{k} \in \mathcal{K}_n\}$, where $R_n(\boldsymbol{k}) = R_n \cap \{\boldsymbol{k}b_n + [0,1)^d b_n\}$ for $\boldsymbol{k} \in \mathcal{K}_n$. That is, we have

$$R_n = \bigcup_{\boldsymbol{k} \in \mathcal{K}_n} R_n(\boldsymbol{k}).$$

Notice that the shape of $R_n(\boldsymbol{k})$ may vary for $\boldsymbol{k} \in \mathcal{K}_{2n}$.

Let $l_n = \{\boldsymbol{l} \in \mathbb{Z}^d : (\boldsymbol{l} + [0,1)^d b_n) \subset R_n\}$ be the index set of the hypercube $(\boldsymbol{l} + [0,1)^d b_n)$ that lies in R_n. Denote $\{I_{\boldsymbol{k}} : \boldsymbol{k} \in \mathcal{K}_n\}$ to be an index set of independent and identically distributed random variables with

$$P_*(I_{\boldsymbol{k}} = \boldsymbol{l}) = \frac{1}{|l_n|} \tag{3.9}$$

for $\boldsymbol{l} \in l_n$, where P_* is the conditional distribution for the resampling method given $\mathcal{S}_n$ and $Y(\boldsymbol{s}_i)$ for $\boldsymbol{s}_i \in \mathcal{S}_n$. Denote $B_n(\boldsymbol{l};\boldsymbol{k}) = R_n(\boldsymbol{k}) - \boldsymbol{k}b_n + \boldsymbol{l}$, where $\boldsymbol{k} \in \mathcal{K}_n$ and $\boldsymbol{l} \in l_n$. Thus, $B_n(\boldsymbol{l};\boldsymbol{k})$ is congruent with $R_n(\boldsymbol{k})$.

we briefly review the resampling method discussed by Lahiri and Zhu (2006) under their setting. Denote $\mathcal{D}_n(R_n) = \{(\boldsymbol{w}(\boldsymbol{s}_i), Y(\boldsymbol{s}_i)) : i = 1, \ldots, n\}$ to be the original observations, and the resampled one to be

$$\mathcal{D}_n^*(R_n) = \{(\boldsymbol{w}(\boldsymbol{s}_i^*), Y(\boldsymbol{s}_i^*)) : \boldsymbol{s}_i^* \in \cup_{\boldsymbol{k} \in \mathcal{K}_n} B_n(I_{\boldsymbol{k}};\boldsymbol{k})\}. \tag{3.10}$$

Let n^* be the sample size of the resampled observations, and it may differ from n by the resampling method. The resampled version of $\hat{\boldsymbol{\beta}}_n$, denoted as $\boldsymbol{\beta}_n^*$, is obtained by solving

$$\sum_{\boldsymbol{k}\in\mathcal{K}_n}\{S_n^*(\boldsymbol{k};\boldsymbol{x}) - \hat{c}_n(\boldsymbol{k})\} = \mathbf{0} \tag{3.11}$$

with respect to $\boldsymbol{x}\in\mathbb{R}^p$, where $S_n^*(\boldsymbol{k};\boldsymbol{x}) = \sum_{i=1}^{n^*}\boldsymbol{w}(\boldsymbol{s}_i^*)\Psi\{Y(\boldsymbol{s}_i^*) - \boldsymbol{w}(\boldsymbol{s}_i^*)^{\mathrm{T}}\boldsymbol{x}\}\mathbb{1}\{\boldsymbol{s}_i^*\in B_n(I_{\boldsymbol{k}};\boldsymbol{k})\}$, $\hat{c}_n(\boldsymbol{k}) = E_*\left[\sum_{i=1}^{n^*}\boldsymbol{w}(\boldsymbol{s}_i^*)\Psi\{\hat{Z}(\boldsymbol{s}_i^*)\}\mathbb{1}(\boldsymbol{s}_i^*\in B_n(I_{\boldsymbol{k}};\boldsymbol{k}))\right]$, $\hat{Z}_n(\boldsymbol{s}_i) = Y(\boldsymbol{s}_i) - \boldsymbol{w}(\boldsymbol{s}_i)^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n$, and E_* is the conditional expectation with respect to the distribution P_* in (3.9). The calibration factor $\hat{c}_n(\boldsymbol{k})$ guarantees that $\boldsymbol{\beta}_n^*$ is an unbiased estimator of $\hat{\boldsymbol{\beta}}_n$ under the conditional distribution P_*.

Denote $P_{\cdot|\boldsymbol{S}}$ to be the conditional probability given $\mathcal{S}_n = \{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n\}$, and we can define $E_{\cdot|\boldsymbol{S}}$ and $V_{\cdot|\boldsymbol{S}}$ for the conditional mean and variance similarly. For the resampling method, we have the following result under the proposed design.

Theorem 3.5. *Suppose that (C1) to (C9) hold and*

$$b_n^{-1} + b_n/\lambda_n = o(1) \qquad \text{as } n\to\infty. \tag{3.12}$$

Then,

$$\sup_{B\in\mathcal{C}}\left|P_*(T_{1n}^*\in B) - P_{\cdot|\boldsymbol{S}}(T_{1n}\in B)\right| \to 0 \qquad \text{in } P_{\cdot|\boldsymbol{S}}\text{-probability}, \ \text{a.s. } (P_{\boldsymbol{S}}), \tag{3.13}$$

where $\mathcal{C} = \mathcal{B}(\mathbb{R}^p)$, $T_{1n}^* = \Lambda_n^{-1}(\boldsymbol{\beta}_n^* - \hat{\boldsymbol{\beta}}_n)$ *and* $T_{1n} = \Lambda_n^{-1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$.

The results in Theorem 3.5 shows that $\boldsymbol{\beta}_n^*$ can be used to make inference for $\boldsymbol{\beta}$ under the proposed design. Thus, a straightforward corollary about the variance estimator of $\hat{\boldsymbol{\beta}}_n$ is given below.

Corollary 3.6. *Suppose that the conditions in Theorem 3.5 hold. Then,*

$$V_*(\boldsymbol{\beta}_n^*) \to V_{\cdot|\boldsymbol{S}}(\hat{\boldsymbol{\beta}}_n) \qquad \text{in } P_{\cdot|\boldsymbol{S}}\text{-probability}, \ \text{a.s. } (P_{\boldsymbol{S}}). \tag{3.14}$$

Corollary 3.6 shows that the variance of $\hat{\boldsymbol{\beta}}$ can be consistently estimated by using $\boldsymbol{\beta}_n^*$ conditional on $\mathcal{S}_n$ almost surely ($P_{\boldsymbol{S}}$). However, the choice of the block size b_n is still an open problem.

We use the method discussed by Hall et al. (1995) to choose the optimal block size. Denote $B_n = \{b_{n,1},\ldots,b_{n,K}\}$ to be a set of valid block sizes satisfying (3.12), where $K\geq 1$. Let

$\{R_n^{(h)} : h = 1, \ldots, H\}$ be a set of pairwise distinct subregions of R_n. For each $b_n \in B_n$, let $b_n^{(h)} = b_n\{\text{vol.}(R_n^{(h)})/\text{vol.}(R_n)\}^{1/(2d)}$ for $h = 1, \ldots, H$. Based on $R_n^{(h)}$ and $b_n^{(h)}$, obtain the variance estimator of $\hat{\boldsymbol{\beta}}_n^{(h)}$ by the resampling method, say $V_n^{*(h)}$, where $\hat{\boldsymbol{\beta}}_n^{(h)}$ solves (3.3) using the observations in $R_n^{(h)}$. The optimal block size is chosen to be the one that minimizes $\sum_{i=1}^{p}\sum_{h=1}^{H}(V_{n,i}^{*(h)} - V_{n,i}^{*})^2$, where V_n^* is the variance estimator of $\hat{\boldsymbol{\beta}}_n$ by the resampling method using the block size b_n and the original observations, and $V_{n,i}^{*(h)}$ and $V_{n,i}^{*}$ are the i-th diagonal element of $V_n^{*(h)}$ and V_n^*, respectively.

3.6 Simulation study

In this section, two simulations are conducted. The first one compares the spatial balance of the proposed design with GRTS and LPM. The second one tests the performance of the M-estimator and the resampling method under the proposed design.

3.6.1 Spatial balance test

Since GRTS and LPM are based on the assumption that the population is finite, we generate 100×100 equally spaced points on the unit square $[0, 1] \times [0, 1]$ as the population, and the first-order inclusion probability is set to be the same for each point. For the proposed design, the sampling region is evenly partitioned, and a uniform distribution is used in each partition region. Three designs are conducted to obtain a sample of size n, and we consider $n \in \{25, 100, 400\}$.

We modify the Voronoi polygon method (Stevens and Olsen, 2004) to measure the spatial balance of a given sample. For a sampled location $\boldsymbol{s}_i$, the Voronoi polygon associated with $\boldsymbol{s}_i$, say V_i, is the set of points that are closer to $\boldsymbol{s}_i$ than other sampled elements, and denote $a_i = \text{vol.}(V_i)$. If the sample is spatially balanced, we expect that $na_i \approx 1$ for $i = 1, \ldots, n$. Thus, we use

$$\zeta = \frac{1}{n}\sum_{i=1}^{n}(na_i - 1)^2$$

to measure the spatial balance for a given sample. Denote ζ_{one}, ζ_{grts} and ζ_{lpm} to be the spatial balance measure for the proposed design, GRTS and LPM respectively. For simplicity, we use statistics $\eta_{one} = \zeta_{one}/\zeta_{grts}$ and $\eta_{lpm} = \zeta_{lpm}/\zeta_{grts}$, and Grafström et al. (2012) showed that $\eta_{lpm} < 1$.

Table 3.1: Monte Carlo mean (outside of the parenthesis) and standard error(inside of the parenthesis) of the spatial balance statistics η. η_{one} is the spatial balance measure for the proposed design against GRTS, and η_{one} is that for LPM against GRTS.

Sample size	η_{one}	η_{lpm}
$n = 25$	0·896(0·069)	0·887(0·078)
$n = 100$	0·748(0·017)	0·701(0·016)
$n = 400$	0·716(0·005)	0·645(0·004)

We conduct 1,000 Monte Carlo simulations for each sample size and design, and Table 3.1 shows the Monte Carlo mean and standard error of statistics η_{one} and η_{lpm}. Compared with GRTS, the sample from the proposed design is more spatially balanced when the sample size is larger. Even though the sample generated by the proposed design is not as spatially balanced as that by LPM, we can modify the sampling density $f_i(\boldsymbol{s})$ to get a sample that spread over the sampling region better.

Remark 1. *As noted by Grafström et al. (2012), the expected computation complexity for LPM is $O(N^2)$, where N is size of the finite population. The procedure of GRTS is so complex that we could not track its expected number of computation, but it is slower than LPM based on our simulation. It can be easily seen that, once the partition is fixed, the computation complexity for the proposed method is $O(n)$, which is much smaller than its two competitors when the size of finite population is large. Furthermore, since the sample from the proposed design is independently obtained in each partition, the proposed method can be accelerated by parallelization, but LPM cannot. Another advantage of the proposed sampling design is that it can generate sample based on an infinite population, but LPM is limited to a finite population.*

3.6.2 Spatial linear regression

In this simulation, the prototype area is set to be $R_0 = (-1/2, 1/2] \times (-1/2, 1/2]$. The population model is

$$Y(\boldsymbol{s}) = \beta_0 + \beta_1 \log(1 + |s_1|) + Z(\boldsymbol{s}), \boldsymbol{s} \in R_n, \tag{3.15}$$

Table 3.2: Summary statistic for comparing the relative efficiency of the M-estimator based on the sample generated by the proposed design and the stochastic design discussed by Lahiri and Zhu (2006).

Dependence	$n = 100$		$n = 400$	
	$eff(\beta_0)$	$eff(\beta_1)$	$eff(\beta_0)$	$eff(\beta_1)$
$r = 1$	0·796	0·774	0·666	0·576
$r = 3$	0·816	0·835	0·747	0·625

where $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_0, \beta_1) = (10, 1)$, and $Z(\boldsymbol{s})$ is a stationary process with spherical semivariogram that has unit sill and range r; see Cressie (2015) for more details. We consider $r \in \{1, 3\}$, $n \in \{400, 900\}$, and set the sampling rate to be $n/\lambda_n^2 = 25/36$. For the proposed design, we partition the sampling region evenly, and use a uniform distribution to draw sample in each partition region.

First, we compare the relative efficiency of the proposed design with the stochastic design discussed by Lahiri and Zhu (2006), and a uniform distribution is used for the latter. Denote $eff(\beta_0) = V_{one}(\hat{\beta}_{n,0})/V_{iid}(\hat{\beta}_{n,0})$, where $\hat{\boldsymbol{\beta}}_n^{\mathrm{T}} = (\hat{\beta}_{n,0}, \hat{\beta}_{n,1})$ can be obtained by solving (3.3), and $V_{one}(\hat{\beta}_{n,0})$ and $V_{iid}(\hat{\beta}_{n,0})$ are the variances of $\hat{\beta}_{n,0}$ under each design, respectively. We use Monte Carlo simulations to estimate $eff(\beta_0)$, and we can estimate $eff(\beta_1)$ in a similar way.

We conduct 1000 Monte Carlo simulations for each design and scenario, and Table 3.2 shows the comparison results. Values of the relative efficiency are less than one, indicating that we gain efficiency by using the proposed design. Besides, as the sample size increases, the proposed design is more efficient than the stochastic design considered by Lahiri and Zhu (2006).

Next, we generate 1000 Monte Carlo samples to obtain the M-estimator of $\boldsymbol{\beta}$, and 2000 resamplings are conducted for each sample. The valid block size set is $B_{n,1} = \{2, 3, 4\}$ when $\lambda_n = 24$, and $B_{n,2} = \{3, 4, 6\}$ when $\lambda_n = 36$. The subregions are chosen to be the four halves of the original sampling region for choosing optimal block size. To test the performance of the resampling method, we consider the square root of mean square error (RMSE) and the relative bias (RB) for the variance estimator, and Table 3.3 summarizes the results. As the sample size increases, RMSE and the absolute value of RB for the variance estimator decrease. For a fixed sample size and block size, RMSE and the absolute value of RB increases as the spatial dependence becomes stronger since the

Table 3.3: Summary statistics for the the variance estimator of β_0 and β_1 by the resampling approach under the proposed sampling design for different scenarios. "RMES" is short for the square root of the mean square error, and "RB" for relative bias. The selected optimal block size is denoted by "†".

Statistics	Dependence	$n = 400$			$n = 900$		
		b_n	β_0	β_1	b_n	β_0	β_1
RMSE	r=1	2†	0·01	0·00	2	N/A	N/A
		3	0·01	0·00	3†	0·00	0·00
		4	0·01	0·00	4	0·00	0·00
		6	N/A	N/A	6	0·00	0·00
	r=3	2†	0·05	0·01	2	N/A	N/A
		3	0·04	0·01	3†	0·02	0·00
		4	0·04	0·01	4	0·02	0·00
		6	N/A	N/A	6	0·02	0·00
RB	r=1	2†	-0·11	-0·06	2	N/A	N/A
		3	-0·15	-0·06	3†	-0·07	-0·02
		4	-0·20	-0·07	4	-0·09	-0·02
		6	N/A	N/A	6	-0·15	-0·03
	r=3	2†	-0·55	-0·51	2	N/A	N/A
		3	-0·45	-0·38	3†	-0·42	-0·37
		4	-0·42	-0·31	4	-0·36	-0·29
		6	N/A	N/A	6	-0·35	-0·23

equivalent number of independent observations decreases; see Cressie (2015) for details. For both of the two cases, the MSE decreases as the sample size increase. We also consider the coverage rate of the 90% confidence interval constructed by the resampling method under the propose design, and Table 3.4 shows the simulation results. As the spatial domain increases, the coverage rate gets closer to 90%. From table 3.3 and table 3.4, we can see that the selected optimal block sizes are reasonable. Besides, for most cases, the mean square errors and the relative bias for the variance estimate based on the optimal block size are smaller than those from other block sizes, and the associated coverage rate under the selected optimal block size is closer to 90%.

Table 3.4: Coverage rate for the 90% confidence interval of β_0 and β_1 by the resampling method under different scenarios. The selected optimal block size is denoted by "†".

Dependence	$n = 100$			$n = 400$		
	b_n	β_0	β_1	b_n	β_0	β_1
r=1	$2^{\dagger}$	0·88	0·89	2	N/A	N/A
	3	0·87	0·89	$3^{\dagger}$	0·89	0·90
	4	0·86	0·89	4	0·88	0·89
	6	N/A	N/A	6	0·87	0·89
r=3	$2^{\dagger}$	0·72	0·74	2	N/A	N/A
	3	0·77	0·80	$3^{\dagger}$	0·79	0·82
	4	0·78	0·82	4	0·81	0·84
	6	N/A	N/A	6	0·82	0·86

3.7 Appendix

Recall that $E_{\boldsymbol{S}}(\cdot)$ is the mean with respect to the proposed one-per-stratum sampling design, $E_{\cdot|\boldsymbol{S}}(\cdot)$ is the conditional mean with respect to the stationary process given the sample $\mathcal{S}_n = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$, and $E_*(\cdot)$ is the conditional mean of the resampling method given the sample $\mathcal{S}_n$ and the corresponding observations $\{Y(\boldsymbol{s}_i) : \boldsymbol{s}_i \in \mathcal{S}_n\}$.

Lemma 3.7. *Suppose (C2) to (C4) hold. Then, for any $\boldsymbol{a} \in \mathbb{R}^d$ with $\|\boldsymbol{a}\| = 1$,*

$$\sigma^2_{n,\boldsymbol{a}} \to \boldsymbol{a}^{\mathrm{T}} H \sigma_\Psi(\boldsymbol{0})\boldsymbol{a} + \boldsymbol{a}^{\mathrm{T}} \left\{ \int Q(\boldsymbol{h})\sigma_\Phi(\boldsymbol{h})\, d\boldsymbol{h} \right\} \boldsymbol{a} \qquad a.s.\ (P_{\boldsymbol{S}}),$$

where $\sigma^2_{n,\boldsymbol{a}} = \sum_{i=1}^n \sum_{j=1}^n d_n(\boldsymbol{S}_i) d_n(\boldsymbol{S}_j) \sigma_\Psi(\boldsymbol{S}_i - \boldsymbol{S}_j)$, $d_n(\boldsymbol{s}) = \boldsymbol{a}^{\mathrm{T}} \Lambda_n^{-1} \boldsymbol{w}(\boldsymbol{s})$, $\boldsymbol{w}(\boldsymbol{s}) = 0$ if $\boldsymbol{s} \notin R_n$, and $\boldsymbol{S}_i$ is the random variable with respect to its realization $\boldsymbol{s}_i$.

Proof of Lemma 3.7. Denote $C(\cdot)$ to be a function of its argument only.

Based on (C8), $d \geq 2$ and Lemma 1.3 discussed by Ibragimov (1962), we can show that $\sigma_\Psi(\boldsymbol{h}) = o(\|\boldsymbol{h}\|^{-3/2})$. Thus, we have

$$\int |\sigma_\Psi(\boldsymbol{h})|^{2r+2}\, \mathrm{d}\boldsymbol{h} < \infty, \tag{3.16}$$

where $r \in \mathbb{N}$ and $\mathbb{N}$ is the set of natural integers. Denote $C_\sigma = \int |\sigma_\Psi(\boldsymbol{h})| \mathrm{d}\boldsymbol{h}$, $C_{2\sigma} = \int |\sigma_\Psi(\boldsymbol{h})|^2\, \mathrm{d}\boldsymbol{h}$, and $C_{4\sigma} = \int |\sigma_\Psi(\boldsymbol{h})|^4\, \mathrm{d}\boldsymbol{h}$.

For simplicity, denote $h_n(\boldsymbol{x},\boldsymbol{y}) = d_n(\boldsymbol{x})d_n(\boldsymbol{y})\sigma_\Psi(\boldsymbol{x}-\boldsymbol{y})$. Thus, we have $\sigma^2_{n,\boldsymbol{a}} = \sum_{i=1}^n \sum_{j=1}^n h_n(\boldsymbol{S}_i,\boldsymbol{S}_j)$. The expectation of $\sigma^2_{n,\boldsymbol{a}}$ with respect to the stochastic sampling procedure is

$$E_{\boldsymbol{S}}(\sigma^2_{n,\boldsymbol{a}}) = E_{\boldsymbol{S}}\left\{\sum_{i=1}^n h_n(\boldsymbol{S}_i,\boldsymbol{S}_i)\right\} + E_{\boldsymbol{S}}\left\{\sum_{i=1}^n \sum_{j\neq i} h_n(\boldsymbol{S}_i,\boldsymbol{S}_j)\right\}.$$

The first part of $E_{\boldsymbol{S}}(\sigma^2_{n,\boldsymbol{a}})$ is

$$\begin{aligned} E_{\boldsymbol{S}}\left\{\sum_{i=1}^n h_n(\boldsymbol{S}_i,\boldsymbol{S}_i)\right\} &= \sigma_\Psi(\boldsymbol{0})\sum_{i=1}^n \int d_n(\boldsymbol{s})^2 f_i(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\ &\to \boldsymbol{a}^{\mathrm{T}} H\boldsymbol{a}\sigma_\Psi(\boldsymbol{0}), \end{aligned} \tag{3.17}$$

where the last equality holds based on (C5). The second part of $E_{\boldsymbol{S}}(\sigma^2_{n,\boldsymbol{a}})$ is

$$\begin{aligned} & E_{\boldsymbol{S}}\left\{\sum_{i=1}^n \sum_{j\neq i} h_n(\boldsymbol{S}_i,\boldsymbol{S}_j)\right\} \\ = & \sum_{i=1}^n \sum_{j\neq i} \int\int d_n(\boldsymbol{x})d_n(\boldsymbol{y})\sigma_\Psi(\boldsymbol{x}-\boldsymbol{y})f_i(\boldsymbol{x})f_j(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \\ = & \sum_{i=1}^n \sum_{j\neq i} \int \sigma_\Psi(\boldsymbol{h})\int d_n(\boldsymbol{y}+\boldsymbol{h})d_n(\boldsymbol{y})f_i(\boldsymbol{y}+\boldsymbol{h})f_j(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{h}, \end{aligned}$$

where the second inequality holds by (C7), $\sigma_\Psi(\boldsymbol{h}) = o(\|\boldsymbol{h}\|^{-3/2})$ and the Fubini's Theorem (Athreya and Lahiri, 2006, Theorem 5.2.2).

Denote $Q_1(\boldsymbol{h}) = \boldsymbol{a}^{\mathrm{T}}Q(\boldsymbol{h})\boldsymbol{a}$, and we have

$$\sum_{i=1}^n \sum_{j\neq i} \int d_n(\boldsymbol{y}+\boldsymbol{h})d_n(\boldsymbol{y})f_i(\boldsymbol{y}+\boldsymbol{h})f_j(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \to Q_1(\boldsymbol{h}) \tag{3.18}$$

as $n\to\infty$ by (C5).

Next, we show that the left part of (3.18) is bounded by a constant for $\boldsymbol{h}\in\mathbb{R}^d$. Consider

$$\begin{aligned} & \left|\sum_{i=1}^n \sum_{j\neq i} \int d_n(\boldsymbol{y}+\boldsymbol{h})d_n(\boldsymbol{y})f_i(\boldsymbol{y}+\boldsymbol{h})f_j(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\right| \\ \leq & \sum_{i=1}^n \sum_{j\neq i} \int_{A_j} |d_n(\boldsymbol{y}+\boldsymbol{h})\mathbb{1}(\boldsymbol{y}+\boldsymbol{h}\in A_i)d_n(\boldsymbol{y})|\, f_i(\boldsymbol{y}+\boldsymbol{h})f_j(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\ \leq & \frac{1}{2}\sum_{i=1}^n \sum_{j\neq i}\left\{\int_{A_j} d_n^2(\boldsymbol{y}+\boldsymbol{h})\mathbb{1}(\boldsymbol{y}+\boldsymbol{h}\in A_i)f_i(\boldsymbol{y}+\boldsymbol{h})^2\mathrm{d}\boldsymbol{y} + \int_{A_j} d_n^2(\boldsymbol{y})\mathbb{1}(\boldsymbol{y}+\boldsymbol{h}\in A_i)f_j(\boldsymbol{y})^2\mathrm{d}\boldsymbol{y}\right\}. \end{aligned} \tag{3.19}$$

In addition to (3.19), we need to bound the following two parts as well before moving on. That is,

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j\neq i}\int_{A_j} d_n^2(\boldsymbol{y}+\boldsymbol{h})\mathbb{1}(\boldsymbol{y}+\boldsymbol{h}\in A_i)f_i(\boldsymbol{y}+\boldsymbol{h})^2\mathrm{d}\boldsymbol{y}\\
\leq\ & M_f\sum_{i=1}^{n}\sum_{j\neq i}\int_{A_j} d_n^2(\boldsymbol{y}+\boldsymbol{h})\mathbb{1}(\boldsymbol{y}+\boldsymbol{h}\in A_i)f_i(\boldsymbol{y}+\boldsymbol{h})\mathrm{d}\boldsymbol{y}\\
=\ & M_f\sum_{i=1}^{n}\int_{\{(R_n\setminus A_i)+\boldsymbol{h}\}\cap A_i} d_n^2(\boldsymbol{y})f_i(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\\
\leq\ & M_f\sum_{i=1}^{n}\int_{A_i} d_n^2(\boldsymbol{y})f_i(\boldsymbol{y})\mathrm{d}\boldsymbol{y},
\end{aligned}\tag{3.20}$$

and

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j\neq i}\int_{A_j} d_n^2(\boldsymbol{y})\mathbb{1}(\boldsymbol{y}+\boldsymbol{h}\in A_i)f_j(\boldsymbol{y})^2\mathrm{d}\boldsymbol{y}\\
\leq\ & M_f\sum_{j=1}^{n}\sum_{i\neq j}\int_{A_j} d_n^2(\boldsymbol{y})\mathbb{1}(\boldsymbol{y}+\boldsymbol{h}\in A_i)f_j(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\\
=\ & M_f\sum_{j=1}^{n}\int_{\{(R_n\setminus A_j)-\boldsymbol{h}\}\cap A_j} d_n^2(\boldsymbol{y})f_j(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\\
\leq\ & M_f\sum_{j=1}^{n}\int_{A_j} d_n^2(\boldsymbol{y})f_j(\boldsymbol{y})\mathrm{d}\boldsymbol{y}
\end{aligned}\tag{3.21}$$

By (3.17) and (3.19) to (3.21), we know that the left part of (3.18) is bounded by a constant, say C_0, when n is sufficiently large.

Thus, by fact that $|Q_1(\boldsymbol{h})|$ is dominated by a constant and $\int|\sigma_\Phi(\boldsymbol{h})|\mathrm{d}\boldsymbol{h}<\infty$, we have

$$E_{\boldsymbol{S}}\left\{\sum_{i=1}^{n}\sum_{j\neq i}h_n(\boldsymbol{S}_i,\boldsymbol{S}_j)\right\}=\int\sigma_\Psi(\boldsymbol{h})Q_1(\boldsymbol{h})\mathrm{d}\boldsymbol{h}\tag{3.22}$$

based on the dominated convergence theorem (Athreya and Lahiri, 2006, Corollary 2.3.13).

By (3.17) and (3.22), we have

$$E_{\boldsymbol{S}}(\sigma_{n,\boldsymbol{a}}^2)=\boldsymbol{a}^{\mathrm{T}}H\sigma_\Psi(\boldsymbol{0})\boldsymbol{a}+\boldsymbol{a}^{\mathrm{T}}\left\{\int\sigma_\Psi(\boldsymbol{h})Q_0(\boldsymbol{h})\mathrm{d}\boldsymbol{h}\right\}\boldsymbol{a}.\tag{3.23}$$

Denote $m_{0n,\boldsymbol{a}}^2=\sup\{\left|\boldsymbol{a}^{\mathrm{T}}\Lambda_n^{-1}\boldsymbol{w}(\boldsymbol{s})\right|^2:\boldsymbol{s}\in\mathbb{R}^d\}$. By $\|\boldsymbol{a}\|=1$, (C7), and the Hölder's inequality (Athreya and Lahiri, 2006, Theorem 3.1.11), we have

$$m_{0n,\boldsymbol{a}}^2=o(n^{-3/4}).\tag{3.24}$$

Now, we consider $E_{\boldsymbol{S}}(\sigma^2_{n,\boldsymbol{a}} - E_{\boldsymbol{S}}\sigma^2_{n,\boldsymbol{a}})^4$. Denote $D_{1n} = \sum_{i=1}^{n}[h_n(\boldsymbol{S}_i, \boldsymbol{S}_i) - E_{\boldsymbol{S}}\{h_n(\boldsymbol{S}_i, \boldsymbol{S}_i)\}]$, $D_{2n} = \sum_{j=1}^{n-1}\sum_{i=j+1}^{n}[h_{1n}^{(i)}(\boldsymbol{S}_j) - E_{\boldsymbol{S}}\{h_n(\boldsymbol{S}_i, \boldsymbol{S}_j)\}]$, $D_{3n} = \sum_{i=2}^{n} U_i$, $U_i = \sum_{j=1}^{i-1}\{h_n(\boldsymbol{S}_i, \boldsymbol{S}_j) - h_{1n}^{(i)}(\boldsymbol{S}_j)\}$, $h_{1n}^{(i)}(\boldsymbol{S}_j) = E_{\boldsymbol{S}}\{h_n(\boldsymbol{S}_i, \boldsymbol{S}_j)|\boldsymbol{S}_j\}$.

Since $h_n(\boldsymbol{x}, \boldsymbol{y}) = h_n(\boldsymbol{y}, \boldsymbol{x})$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, we have

$$\sigma^2_{n,\boldsymbol{a}} - E_{\boldsymbol{S}}(\sigma^2_{n,\boldsymbol{a}}) = D_{1n} + 2D_{2n} + 2D_{3n}. \tag{3.25}$$

Before proceeding, for $r \in \mathbb{N}_+$ and $i = 1, \ldots, n$, consider

$$\begin{aligned} E_{\boldsymbol{S}}\{h_n^{2r}(\boldsymbol{S}_i, \boldsymbol{S}_i)\} &= \int d_n^{4r}(\boldsymbol{s})\sigma_\Psi^{2r}(\mathbf{0})f_i(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\ &\leq M_f\sigma_\Psi^{2r}(\mathbf{0})m_{0n,\boldsymbol{a}}^{4r}M_A, \qquad (3.26) \\ \sum_{j\in J} E_{\boldsymbol{S}}\{h_n^{2r}(\boldsymbol{S}_i, \boldsymbol{S}_j)|\boldsymbol{S}_i\} &= \sum_{j\in J}\int_{A_j} d_n^{2r}(\boldsymbol{s})d_n^{2r}(\boldsymbol{S}_i)\sigma_\Psi^{2r}(\boldsymbol{s} - \boldsymbol{S}_i)f_j(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\ &\leq M_f m_{0n,\boldsymbol{a}}^{4r}\int |\sigma_\Psi(\boldsymbol{s})|^{2r}\,\mathrm{d}\boldsymbol{s}, \qquad (3.27) \end{aligned}$$

where J is a subset of $\{1, \ldots, n\} \setminus \{i\}$ in (3.27), and recall that $f_i(\boldsymbol{s})$ is zero outside of A_i. For $j = 1, \ldots, n-1$, consider

$$\sum_{i=j+1}^{n} h_n^{(i)}(\boldsymbol{S}_j) \leq m_{0n,\boldsymbol{a}}^2 M_f \int_{\cup_{i=j+1}^{n}A_i}|\sigma_\Psi(\boldsymbol{s} - \boldsymbol{S}_j)|\mathrm{d}\boldsymbol{s} \leq m_{0n,\boldsymbol{a}}^2 M_f C_\sigma,$$
$$E_{\boldsymbol{S}}\left(\sum_{i=j+1}^{n}\left[h_n^{(i)}(\boldsymbol{S}_j) - E_{\boldsymbol{S}}\{h_n^{(i)}(\boldsymbol{S}_j)\}\right]\right)^{2r} \leq CE_{\boldsymbol{S}}\left(\sum_{i=j+1}^{n}\left[h_n^{(i)}(\boldsymbol{S}_j)\right]\right)^{2r} \leq CC_\sigma^{2r}M_f^{2r}m_{0n,\boldsymbol{a}}^{4r}.$$

For D_{1n}, it is a summation of n independent random variables with mean zero. Thus, we have

$$\begin{aligned} E_{\boldsymbol{S}}(D_{1n}^4) &\leq C\left\{\sum_{i=1}^{n} E_{\boldsymbol{S}}h_n^4(\boldsymbol{S}_i, \boldsymbol{S}_i) + \sum_{i=1}^{n}\sum_{j\neq i} E_{\boldsymbol{S}}h_n^2(\boldsymbol{S}_i, \boldsymbol{S}_i)E_{\boldsymbol{S}}h_n^2(\boldsymbol{S}_j, \boldsymbol{S}_j)\right\} \\ &\leq C\{m_{0n,\boldsymbol{a}}^8 M_f M_A\sigma_\Psi^4(\mathbf{0})n + m_{0n,\boldsymbol{a}}^8\sigma_\Phi^4(\mathbf{0})M_f^2M_A^2n^2\} \\ &\leq C\{M_f, M_A, \sigma_\Psi(\mathbf{0})\}m_{0n,\boldsymbol{a}}^8 n^2, \qquad (3.28) \end{aligned}$$

where the second inequality holds by (3.26), and recall that $C\{M_f, M_A, \sigma_\Psi(\mathbf{0})\}$ is a function of M_f, M_A, and $\sigma_\Psi(\mathbf{0})$.

We can use a similar technique for bounding $E_{\boldsymbol{S}}(D_{1n}^4)$ to obtain the upper bound of $E_{\boldsymbol{S}}(D_{2n}^4)$ as follows. That is,

$$\begin{aligned} E_{\boldsymbol{S}}(D_{2n}^4) &\leq C\left[\sum_{j=1}^{n-1} E_{\boldsymbol{S}}\left(\sum_{i=j+1}^{n}\left[h_n^{(i)}(\boldsymbol{S}_j) - E_{\boldsymbol{S}}\{h_n^{(i)}(\boldsymbol{S}_j)\}\right]\right)^4 + \right. \\ &\left. \sum_{j=1}^{n-1}\sum_{k\neq j} E_{\boldsymbol{S}}\left(\sum_{i=j+1}^{n}\left[h_n^{(i)}(\boldsymbol{S}_j) - E_{\boldsymbol{S}}\{h_n^{(i)}(\boldsymbol{S}_j)\}\right]\right)^2 E_{\boldsymbol{S}}\left(\sum_{i=k+1}^{n}\left[h_n^{(i)}(\boldsymbol{S}_j) - E_{\boldsymbol{S}}\{h_n^{(i)}(\boldsymbol{S}_j)\}\right]\right)^2\right] \\ &\leq C(C_\sigma^4 M_f^4 m_{0n,\boldsymbol{a}}^8 n + C_\sigma^4 M_f^4 m_{0n,\boldsymbol{a}}^8 n^2) \\ &\leq C(C_\sigma, M_f) m_{0n,\boldsymbol{a}}^8 n^2 \end{aligned} \tag{3.29}$$

Next, we consider D_{3n}. Note the fact that $E_{\boldsymbol{S}}(U_i|\boldsymbol{S}_1,\ldots,\boldsymbol{S}_{i-1}) = 0$ for $i = 2,\ldots,n$. Thus, $\left\{\sum_{j=2}^{i} U_j, \mathcal{F}_i^{\boldsymbol{S}}\right\}_{i=2}^{n}$ is a martingale, where $\mathcal{F}_i^{\boldsymbol{S}} = \sigma\langle \boldsymbol{S}_1,\ldots,\boldsymbol{S}_i\rangle$.

By Rosenthal's inequality (Hall and Heyde, 1980, Theorem 2.12), we have

$$\begin{aligned} E_{\boldsymbol{S}}(D_{3n}^4) &\leq C\left[E_{\boldsymbol{S}}\left\{\sum_{i=2}^{n} E_{\boldsymbol{S}}(U_i^2|\mathcal{F}_{i-1}^{\boldsymbol{S}})\right\}^2 + \sum_{i=2}^{n} E_{\boldsymbol{S}} U_i^4\right] \\ &\leq C\left(E_{\boldsymbol{S}}\left[(n-1)\sum_{i=2}^{n}\{E_{\boldsymbol{S}}(U_i^2|\mathcal{F}_{i-1}^{\boldsymbol{S}})\}^2\right] + \sum_{i=2}^{n} E_{\boldsymbol{S}} U_i^4\right) \\ &\leq Cn\left\{\sum_{i=2}^{n} E_{\boldsymbol{S}} U_i^4\right\} \\ &\leq C_1 n\left\{\sum_{i=2}^{n} E_{\boldsymbol{S}}(E_{\boldsymbol{S}}\left[\{U_i - E_{\boldsymbol{S}}(U_i|\boldsymbol{S}_i)\}^4|\boldsymbol{S}_i\right] + \{E_{\boldsymbol{S}}(U_i|\boldsymbol{S}_i)\}^4)\right\}. \end{aligned} \tag{3.30}$$

Notice that U_i is a sum of $i-1$ independent random variables given $\boldsymbol{S}_i$, so we have

$$\begin{aligned} E_{\boldsymbol{S}}\left[\{U_i - E_{\boldsymbol{S}}(U_i|\boldsymbol{S}_i)\}^4|\boldsymbol{S}_i\right] &\leq C\left[\sum_{j=1}^{i-1} E_{\boldsymbol{S}}\{h_n^4(\boldsymbol{S}_i,\boldsymbol{S}_j) \mid \boldsymbol{S}_i\}\right. \\ &\left. + \sum_{j=1}^{i-1}\sum_{k\neq j} E_{\boldsymbol{S}}\{h_n^2(\boldsymbol{S}_i,\boldsymbol{S}_j) \mid \boldsymbol{S}_i\} E_{\boldsymbol{S}}\{h_n^2(\boldsymbol{S}_i,\boldsymbol{S}_k) \mid \boldsymbol{S}_i\}\right] \\ &\leq C(M_f m_{0n,\boldsymbol{a}}^8 C_{4\sigma} + M_f^2 m_{0n,\boldsymbol{a}}^8 C_{2\sigma}^2) \\ &= C(M_f, C_{2\sigma}, C_{4\sigma}) m_{0n,\boldsymbol{a}}^8 \end{aligned} \tag{3.31}$$

where the second inequality is based on (3.27). Besides, we have

$$\begin{aligned} |E_{\boldsymbol{S}}(U_i|\boldsymbol{S}_i)| &\leq \sum_{j=1}^{i-1} E_{\boldsymbol{S}}\{|h_n(\boldsymbol{S}_i,\boldsymbol{S}_j)| \mid \boldsymbol{S}_i\} + \sum_{j=1}^{i-1} E_{\boldsymbol{S}}\{|h_n(\boldsymbol{S}_i,\boldsymbol{S}_j)|\} \\ &\leq C_\sigma M_f m_{0n,\boldsymbol{a}}^2 + M_A C_\sigma M_f m_{0n,\boldsymbol{a}}^2, \end{aligned}$$

where the first part in the second inequality can be derived by a similar argument in (3.27), and the second part is obtained by (C2) and integration of (3.27) over A_i. Therefore,

$$E_{\boldsymbol{S}}\left\{E_{\boldsymbol{S}}(U_i|\boldsymbol{S}_i)\right\}^4 \leq C(C_\sigma, M_f, M_A)m_{0n,\boldsymbol{a}}^8. \tag{3.32}$$

Thus, by (3.28)–(3.32), we have

$$\begin{aligned}\sum_{n=1}^{\infty} E_{\boldsymbol{S}}\{\sigma_{n,\boldsymbol{a}} - E_{\boldsymbol{S}}(\sigma_{n,\boldsymbol{a}})\}^4 \quad &\leq \quad \sum_{n=1}^{\infty} C(M_f, M_A, \sigma_\Psi(\mathbf{0}), C_\sigma, C_{2\sigma}, C_{4\sigma})n^2 m_{0n,\boldsymbol{a}}^8 \\ &< \quad \infty, \end{aligned} \tag{3.33}$$

where the last equality holds based on (3.24). Therefore, by the Borel-Cantelli Lemma (Athreya and Lahiri, 2006, Theorem 7.2.2) and Markov's inequality (Athreya and Lahiri, 2006, Proposition 6.2.4), we have proved Lemma 3.7. □

Proof of Theorem 3.1. Based on Lemma (C3), Lemma 3.7 and Lemma 1.3 discussed by Ibragimov (1962), we could use a similar blocking argument in Lahiri (2003) to prove this theorem, and we refer readers to Lahiri (2003) for more details. □

Lemma 3.8. *Let $g : R_n \to \mathbb{R}$ be a Borel measurable function satisfying $E[|g\{Z(\mathbf{0})\}|] < \infty$ and $E[g\{Z(\mathbf{0})\}] = 0$ for $i = 1, \ldots, n$. Also, let $a_{in} = a_{in}(\boldsymbol{S}_i)$, $i = 1, \ldots, n$ be $\sigma\langle \boldsymbol{S}_i\rangle$ measurable random variables such that*

$$\sum_{i=1}^{n} |a_{in}(\boldsymbol{S}_i)| = O(1), \qquad a.s. \ (P_{\boldsymbol{S}}), \tag{3.34}$$

and

$$\sum_{i=1}^{n} a_{in}^2(\boldsymbol{S}_i) = o(1), \qquad a.s. \ (P_{\boldsymbol{S}}). \tag{3.35}$$

Then, $\sum_{i=1}^{n} a_{in}(\boldsymbol{S}_i)g(Z(\boldsymbol{S}_i)) \to 0$ in $P_{\cdot|\boldsymbol{S}}$-probability, a.s. ($P_{\boldsymbol{S}}$).

The proof of Lemma 3.8 uses the similar steps as discussed by Lahiri (2003), so we omit the details.

Proof of Theorem 3.3. The proof mainly follows the one in Theorem 3.1 Lahiri and Mukherjee (2004). We only give the proof for the first part, and the proof for the last two parts is the same.

Recall that $E_{\cdot|\boldsymbol{S}}(\cdot)$ is the conditional mean given $\mathcal{S}_n$, and we would like to show that, for any $b \in (0, \infty)$,

$$\sup_{\|\boldsymbol{u}\| \leq b} \left\| \Lambda_n^{-1}\{M_n(\boldsymbol{\beta} + \Lambda_n^{-1}\boldsymbol{u}) - M_n(\boldsymbol{\beta})\} + HE_{\cdot|\boldsymbol{S}}[\Psi'\{Z(\mathbf{0})\}]\boldsymbol{u} \right\| = o_p(1). \tag{3.36}$$

Denote $\boldsymbol{v}_i = \Lambda_n^{-1}\boldsymbol{w}(\boldsymbol{s}_i)$, so we have

$$\begin{aligned}
\Lambda_n^{-1}\{M_n(\boldsymbol{\beta} + \Lambda_n^{-1}\boldsymbol{u}) - M_n(\boldsymbol{\beta})\} &= \Lambda_n^{-1} \sum_{i=1}^n \boldsymbol{w}(\boldsymbol{s})[\Psi\{Z(\boldsymbol{s}_i) - \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{u}\} - \Psi\{Z(\boldsymbol{s}_i)\}] \\
&= \Lambda_n^{-1} \sum_{i=1}^n \boldsymbol{w}(\boldsymbol{s}) \int_{Z(\boldsymbol{s}_i)}^{Z(\boldsymbol{s}_i) - \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{u}} \Psi'(t)\mathrm{d}t \\
&= \Lambda_n^{-1} \sum_{i=1}^n \boldsymbol{w}(\boldsymbol{s}) \int_0^{-\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{u}} \Psi'\{Z(\boldsymbol{s}_i) + t\}\mathrm{d}t.
\end{aligned}$$

Denote $t_i = \sup\{\left|\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{u}\right| : \|\boldsymbol{u}\| \leq b\} \leq b\|\boldsymbol{v}_i\|$, and $t_i = o(1)$ based on (C7). By taking conditional expectation, we have

$$\begin{aligned}
& E_{\cdot|\boldsymbol{S}} \sup_{\|\boldsymbol{u}\| \leq b} \left\| \Lambda_n^{-1}\{M_n(\boldsymbol{\beta} + \Lambda_n^{-1}\boldsymbol{u}) - M_n(\boldsymbol{\beta})\} + \Lambda_n^{-1} \sum_{i=1}^n \boldsymbol{w}(\boldsymbol{s}_i)\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{u}\Psi'(Z(\boldsymbol{s}_i)) \right\| \\
\leq & \sum_{i=1}^n \|\boldsymbol{v}_i\| \int_0^{\left|\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{u}\right|} E_{\cdot|\boldsymbol{S}} \left|\Phi'\{Z(\boldsymbol{s}_i) + t\} - \Psi'\{Z(\boldsymbol{s}_i)\}\right| \mathrm{d}t \\
\leq & \sum_{i=1}^n \|\boldsymbol{v}_i\| \int_0^{t_i} E_{\cdot|\boldsymbol{S}} \left|\Phi'\{Z(\boldsymbol{s}_i) + t\} - \Psi'\{Z(\boldsymbol{s}_i)\}\right| \mathrm{d}t \\
\leq & \frac{1}{(1+\gamma)} \sum_{i=1}^n \|\boldsymbol{v}_i\|^{2+\gamma} \\
= & \ o(1),
\end{aligned} \tag{3.37}$$

where C_γ is a constant, and the third inequality is based on (C9), and the last equality is by (C7).

Based on (C7), we have $\left|\|\boldsymbol{v}_i\|^2 - E_{\boldsymbol{S}}\|\boldsymbol{v}_i\|^2\right| < 2n^{-1/2}$ for $i = 1, \ldots, n$. Therefore, based on Bernstein's inequality (Bennett, 1962), for any $\epsilon > 0$, we have $P_{\boldsymbol{S}}\left(\left|\sum_{i=1}^n\{\|\boldsymbol{v}_i\|^2 - E_{\boldsymbol{S}}\|\boldsymbol{v}_i\|^2\}\right| > \epsilon\right) \leq \exp\left\{-O(n^{1/2})\right\}$, where the last inequality is based on (C7). Thus, by the Borel-Cantelli Lemma, we have

$$\textstyle\sum_{i=1}^n \|\boldsymbol{v}_i\|^2 - E_{\boldsymbol{S}}\sum_{i=1}^n \|\boldsymbol{v}_i\|^2 \to 0 \quad \text{a.s. } (P_{\boldsymbol{S}}), \tag{3.38}$$

$$\textstyle E_{\boldsymbol{S}}\sum_{i=1}^n \|\boldsymbol{v}_i\|^2 = \mathrm{tr}\left(E_{\boldsymbol{S}}\sum_{i=1}^n \|\boldsymbol{v}_i\|^2\right) \to \mathrm{tr}(H), \tag{3.39}$$

where tr$(\cdot)$ is the trace operator, and (3.39) is based on (C5). Based on the results (3.38)–(3.39), we have

$$\sum_{i=1}^{n} \left\| \Lambda_n^{-1} \boldsymbol{w}(\boldsymbol{s}_i) \boldsymbol{v}_i^{\mathrm{T}} \right\| = O(1) \tag{3.40}$$

almost surely.

By noting the fact that $\|\boldsymbol{v}_i\|^4 = o(n^{-1})$ by (C7), we have

$$\sum_{i=1}^{n} \left\| \Lambda_n^{-1} \boldsymbol{w}(\boldsymbol{s}_i) \boldsymbol{v}_i^{\mathrm{T}} \right\|^2 \leq \sum_{i=1}^{n} \|\boldsymbol{v}_i\|^4 = o(1). \tag{3.41}$$

By (3.40), (3.41) and Lemma 3.8, we have

$$\begin{aligned} & \sup_{\|\boldsymbol{u}\| \leq b} \left\| \Lambda_n^{-1} \sum_{i=1}^{n} \boldsymbol{w}(\boldsymbol{s}_i) \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{u} [\Psi'\{Z(\boldsymbol{s}_i)\} - E_{\cdot|\boldsymbol{S}} \Psi'\{Z(\boldsymbol{0})\}] \right\| \\ \leq \quad & b \left\| \Lambda_n^{-1} \sum_{i=1}^{n} \boldsymbol{w}(\boldsymbol{s}_i) \boldsymbol{v}_i^{\mathrm{T}} [\Psi'\{Z(\boldsymbol{s}_i)\} - E_{\cdot|\boldsymbol{S}} \Psi'\{Z(\boldsymbol{0})\}] \right\| = o_p(1), \text{ a.s. } (P_{\boldsymbol{S}}). \end{aligned} \tag{3.42}$$

Based on (3.37) to (3.39), (3.42) and the Markov's inequality, we have

$$\sup_{\|\boldsymbol{u}\| \leq b} \left\| \Lambda_n^{-1} \{M_n(\boldsymbol{\beta} + \Lambda_n^{-1} \boldsymbol{u}) - M_n(\boldsymbol{\beta})\} + H E_{\cdot|\boldsymbol{S}} \Psi'\{Z(\boldsymbol{0})\} \boldsymbol{u} \right\| = o_p(1) \text{ a.s. } (P_{\boldsymbol{S}}) \tag{3.43}$$

for $b \in (0, \infty)$.

The remaining proof is approximately the same with the one shown in Theorem 3.1 discussed by Lahiri (2004). Thus, by Lemma 3.7 and Theorem 3.1, we can show that

$$\boldsymbol{a}^{\mathrm{T}} \Lambda_n^{-1} M_n(\boldsymbol{\beta}) \boldsymbol{a} \xrightarrow{d} N(0, \boldsymbol{a}^{\mathrm{T}} \Sigma_M \boldsymbol{a}), \text{ a.s. } (P_{\boldsymbol{S}}), \tag{3.44}$$

where $\boldsymbol{a} \in \mathbb{R}^p$ with $\|\boldsymbol{a}\| = 1$. Thus, the theorem is proved by Cramer-Wold device. $\square$

Proof of Theorem 3.4. By (C6) to (C7) and $g(\boldsymbol{s}) = \{\text{vol.}(R_0)\}^{-1}$ for $\boldsymbol{s} \in R_0$, we have, by Lahiri and Zhu (2006),

$$\lambda_n^{d/2} \Lambda_{n,1} (\hat{\boldsymbol{\beta}}_{n,iid} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \chi_0^{-2} \Sigma_{\beta,iid}), \tag{3.45}$$

where $\hat{\boldsymbol{\beta}}_{n,iid}$ solves (3.3) based on the independent and identically distributed design associated with $g(\boldsymbol{s})$, and $\Sigma_{\beta,iid} = c^{-1} H_{iid}^{-1} \sigma_\Psi(\boldsymbol{0}) + H_{iid}^{-1} \{\int \sigma_\Psi(\boldsymbol{h}) Q_{iid}(\boldsymbol{h}) \mathrm{d}\boldsymbol{h}\} H_{iid}^{-1}$, and recall that $n/\lambda_n^d \to c \in (0, \infty)$.

Thus, by the fact that $n/\lambda_n^d \to c$, (3.45) and Slutsky's theorem (Athreya and Lahiri, 2006), we have

$$\sqrt{n}\Lambda_{n,1}(\hat{\boldsymbol{\beta}}_{n,iid} - \boldsymbol{\beta}) \xrightarrow{d} N(0, c\chi_0^{-2}\Sigma_{\boldsymbol{\beta},iid}) \quad \text{a.s. } (P_{iid}), \tag{3.46}$$

where P_{iid} is the probability measure for the independent and identically distributed sampling design.

First, we prove that the first asymptotic property in (C5) holds under this one-per-stratum sampling design. Consider

$$\begin{aligned}\sum_{i=1}^{n}\int \boldsymbol{w}(\boldsymbol{s})\boldsymbol{w}(\boldsymbol{s})^{\mathrm{T}} f_i(\boldsymbol{s})\mathrm{d}\boldsymbol{s} &= \frac{n}{\lambda_n^d \mathrm{vol.}(R_0)}\int_{R_n} \boldsymbol{w}(\boldsymbol{s})\boldsymbol{w}(\boldsymbol{s})^{\mathrm{T}}\mathrm{d}\boldsymbol{s} \\ &= \frac{n}{\mathrm{vol.}(R_0)}\int_{R_0} \boldsymbol{w}(\lambda_n\boldsymbol{s})\boldsymbol{w}(\lambda_n\boldsymbol{s})^{\mathrm{T}}\mathrm{d}\boldsymbol{s}, \end{aligned} \tag{3.47}$$

where the second equality is based on the change of integral variable. By (C5), (3.7) and (3.47), we have $\Lambda_n^{-1}\left\{\sum_{i=1}^n \int \boldsymbol{w}(\boldsymbol{s})\boldsymbol{w}(\boldsymbol{s})^{\mathrm{T}} f_i(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right\}\Lambda_n^{-1} \to H_{iid} \quad$ as $n \to \infty$, where $\Lambda_n = \sqrt{n}\Lambda_{n,iid}$.

Next, for $\boldsymbol{h} \in \mathbb{R}^d$, consider

$$\begin{aligned} & \sum_{i=1}^{n}\sum_{j\neq i}\int \boldsymbol{w}(\boldsymbol{y}+\boldsymbol{h})\boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}} f_i(\boldsymbol{y}+\boldsymbol{h})f_j(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\ = & \left\{\frac{n}{\lambda_n^d \mathrm{vol.}(R_0)}\right\}^2 \sum_{i=1}^{n}\sum_{j\neq i}\int_{A_j\cap(A_i-\boldsymbol{h})} \boldsymbol{w}(\boldsymbol{y}+\boldsymbol{h})\boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}}\mathrm{d}\boldsymbol{y} \\ = & \left\{\frac{n}{\lambda_n^d \mathrm{vol.}(R_0)}\right\}^2 \sum_{i=1}^{n}\int_{R_n\cap\{A_i^{\mathrm{C}}\cap(A_i-\boldsymbol{h})\}} \boldsymbol{w}(\boldsymbol{y}+\boldsymbol{h})\boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}}\mathrm{d}\boldsymbol{y} \\ \preceq & \left\{\frac{n}{\lambda_n^d \mathrm{vol.}(R_0)}\right\}^2 \int_{R_n\cap\{\cup_{i=1}^n(A_i-\boldsymbol{h})\}} \boldsymbol{w}(\boldsymbol{y}+\boldsymbol{h})\boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}}\mathrm{d}\boldsymbol{y} \\ \preceq & \left\{\frac{n}{\lambda_n^d \mathrm{vol.}(R_0)}\right\}^2 \int_{R_n} \boldsymbol{w}(\boldsymbol{y}+\boldsymbol{h})\boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}}\mathrm{d}\boldsymbol{y} \\ = & \frac{n^2}{\lambda_n^d}\frac{1}{\{\mathrm{vol.}(R_0)\}^2}\int_{R_0} \boldsymbol{w}(\boldsymbol{y}+\boldsymbol{h})\boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}}\mathrm{d}\boldsymbol{y}, \end{aligned} \tag{3.48}$$

where A^{C} is the complement of set A, and $B \preceq C$ for two matrix if $C - B$ is a non-negative definitive matrix for two square matrices B and C. Thus, by (C5), (3.7) and (3.48), we have, for $\boldsymbol{h} \in \mathbb{R}^d$,

$$Q(\boldsymbol{h}) \preceq cQ_{iid}(\boldsymbol{h}). \tag{3.49}$$

Based on Theorem 3.3, we have

$$\sqrt{n}\Lambda_{n,1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} N(0, c\chi_0^{-2}\Sigma_{\boldsymbol{\beta}}) \quad \text{a.s. } (P_{\boldsymbol{S}}) \tag{3.50}$$

where $\Sigma_{\beta} = H_{iid}^{-1}\sigma_{\Psi}(\mathbf{0}) + H_{iid}^{-1}\{\int \sigma_{\Psi}(\boldsymbol{h})Q(\boldsymbol{h})\mathrm{d}\boldsymbol{h}\}H_{iid}^{-1}$. Thus, the first result of Theorem 3.4 is proved by (3.46), (3.49), and (3.50).

Denote $B_{\boldsymbol{h}} = R_n \cap (R_n - \boldsymbol{h})^{\mathrm{C}}$ for $\boldsymbol{h} \in \mathbb{R}^d$, and it is easy to check that $B_{\boldsymbol{h}}$ has positive Lebesgue measure. Therefore,

$$\int_{R_n} \boldsymbol{w}(\boldsymbol{y}+\boldsymbol{h})\boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}}\mathrm{d}\boldsymbol{y} - \int_{R_n \cap \{\cup_{i=1}^{n}(A_i - \boldsymbol{h})\}} \boldsymbol{w}(\boldsymbol{y}+\boldsymbol{h})\boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}}\mathrm{d}\boldsymbol{y} = \int_{B_{\boldsymbol{h}}} \boldsymbol{w}(\boldsymbol{y}+\boldsymbol{h})\boldsymbol{w}(\boldsymbol{y})^{\mathrm{T}}\mathrm{d}\boldsymbol{y}$$

is positive definitive. Thus, $Q(\boldsymbol{h}) \prec cQ_{iid}(\boldsymbol{h})$, and the second part of theorem 3.4 is proved. □

Lemma 3.9. *Suppose that (C1)–(C9) hold. Then,*

$$\left\|\hat{\Sigma}_n - \Sigma_M\right\| \to 0 \quad \text{in } P_{\cdot|\boldsymbol{S}}\text{-probability, a.s. } (P_{\boldsymbol{S}}), \tag{3.51}$$

where $\hat{\Sigma}_n = \sum_{\boldsymbol{k}\in\mathcal{K}_n} V_*\{\Lambda_n^{-1}S_n^*(\boldsymbol{k},\hat{\boldsymbol{\beta}}_n)\}$, *and recall that* $\Sigma_M = H\sigma_{\Psi}(\mathbf{0}) + \int \sigma_{\Psi}(\boldsymbol{h})Q(\boldsymbol{h})\mathrm{d}\boldsymbol{h}$.

Proof of Lemma 3.9. The argument here is the based on the proof of Lemma 3 of Lahiri and Zhu (2006), and we only consider the case where $p = 1$. For higher dimensional space, similar argument can be made.

Denote $\tilde{S}_n(\boldsymbol{l};\boldsymbol{k}) = \sum_{i=1}^{n}\boldsymbol{w}(\boldsymbol{s}_i)\Psi\{Z(\boldsymbol{s}_i)\}\mathbb{1}(\boldsymbol{s}_i \in B_n(\boldsymbol{l};\boldsymbol{k}))$, where $\boldsymbol{l} \in l_n$ and $\boldsymbol{k} \in \mathcal{K}_n$. Let $\tilde{\Sigma}_n = \sum_{\boldsymbol{k}\in\mathcal{K}_n}(|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}\{\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\}^2 - [|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}\{\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\}]^2)$. Thus, by (C3), (C9) and Theorem 3.3, we have

$$\hat{\Sigma}_n - \tilde{\Sigma}_n \to 0 \quad \text{in } P_{\cdot|\boldsymbol{S}}\text{-probability, a.s. } (P_{\boldsymbol{S}}), \tag{3.52}$$

and recall that $\hat{\Sigma}_n = \sum_{\boldsymbol{k}\in\mathcal{K}_n}(|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}\{\Lambda_n^{-1}\hat{S}_n(\boldsymbol{l};\boldsymbol{k})\}^2 - [|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}\{\Lambda_n^{-1}\hat{S}_n(\boldsymbol{l};\boldsymbol{k})\}]^2)$, where $\hat{S}_n(\boldsymbol{l};\boldsymbol{k}) = \sum_{i=1}^{n}\boldsymbol{w}(\boldsymbol{s}_i)\Psi\{\hat{Z}(\boldsymbol{s}_i)\}\mathbb{1}\{\boldsymbol{s}_i \in B_n(\boldsymbol{l};\boldsymbol{k})\}$.

By Lemma 2 of Lahiri and Zhu (2006) and (C3), we have

$$\begin{aligned}
\sum_{\boldsymbol{k}\in\mathcal{K}_n} E_{\cdot|\boldsymbol{S}}\left(|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}[\{\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\}^2 - E_{\cdot|\boldsymbol{S}}\{\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\}^2]\right)^2 &= o(1), \\
\sum_{\boldsymbol{k}\in\mathcal{K}_n} E_{\cdot|\boldsymbol{S}}\left\{|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\right\}^2 &= o(1).
\end{aligned}$$

In order to conclude the proof, it remains to show the following results of the variance estimator. That is,

$$E_{\boldsymbol{S}}\left[\sum_{\boldsymbol{k}\in\mathcal{K}_n}|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}E_{\cdot|\boldsymbol{S}}\left\{\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\right\}^2\right]\to\Sigma_M \quad \text{as } n\to\infty, \tag{3.53}$$

$$\sum_{\boldsymbol{k}\in\mathcal{K}_n}|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}E_{\cdot|\boldsymbol{S}}\left\{\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\right\}^2\to E_{\boldsymbol{S}}\left[\sum_{\boldsymbol{k}\in\mathcal{K}_n}|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}E_{\cdot|\boldsymbol{S}}\left\{\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\right\}^2\right] \tag{3.54}$$

almost surely. Notice that the proof of (3.54) is similar with the one in Lemma 3.7, so we only show (3.53). Denote $\tilde{\Sigma}_{jn}=\sum_{\boldsymbol{k}\in\mathcal{K}_{jn}}|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}E_{\cdot|\boldsymbol{S}}\left(\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\right)^2$ for $j=1,2$. Then,

$$\begin{aligned}
E_{\boldsymbol{S}}\left[\sum_{\boldsymbol{k}\in\mathcal{K}_{1n}}|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}E_{\cdot|\boldsymbol{S}}\left\{\Lambda_n^{-1}\tilde{S}_n(\boldsymbol{l};\boldsymbol{k})\right\}^2\right] &= |\mathcal{K}_{1n}|\,|l_n|^{-1}\sum_{\boldsymbol{l}\in l_n}\left[\sum_{i=1}^n E_{\boldsymbol{S}}(v_i^2)\sigma_\Psi(\boldsymbol{0})\mathbb{1}\{\boldsymbol{s}_i\in B_n(\boldsymbol{l};\boldsymbol{0})\}\right.\\
&\quad\left.+\sum_{i=1}^n\sum_{j\neq i}E_{\boldsymbol{S}}(v_iv_j)\sigma_\Psi(\boldsymbol{s}_i-\boldsymbol{s}_j)\mathbb{1}\{\boldsymbol{s}_i,\boldsymbol{s}_j\in B_n(\boldsymbol{l};\boldsymbol{0})\}\right]\\
&= \Sigma_{11n}+\Sigma_{12n},\ \text{say}\ .
\end{aligned}$$

Notice that $|\mathcal{K}_{1n}|=\lambda_n^d b_n^{-d}\text{vol.}(R_0)(1+o(1))$ and $|l_n|=\lambda_n^d\text{vol.}(R_0)(1+o(1))$. Denote $R_{2n}=\cup_{\boldsymbol{k}\in(\mathcal{K}_{1n}\cap R_{1n})}R_n(\boldsymbol{k})$, where $R_{1n}=\lambda_n(R_0\backslash R_0^{b_n\lambda_n^{-1}})$. It can be shown that $\left|\{\boldsymbol{l}\in l_n:\boldsymbol{s}\in\boldsymbol{l}+b_n[0,1]^d\}\right|=b_n^d\{1+o(1)\}$ for $\boldsymbol{s}\in R_{2n}$. By (C1) and (C5), we have

$$\begin{aligned}
\Sigma_{11n} &= \Lambda_n^{-1}\frac{|\mathcal{K}_{1n}|}{|l_n|}\sigma_\Psi(\boldsymbol{0})\left[\sum_{i=1}^n\int_{R_{2n}}w_n^2(\boldsymbol{s})f_i(\boldsymbol{s})\sum_{\boldsymbol{l}\in l_n}\mathbb{1}\{\boldsymbol{s}\in B_n(\boldsymbol{l};\boldsymbol{0})\}\mathrm{d}\boldsymbol{s}\right]\Lambda_n^{-1}(1+o(1))\\
&= \sigma_\Psi(\boldsymbol{0})H(1+o(1)).
\end{aligned}\tag{3.55}$$

Similarly, we have

$$\begin{aligned}
\Sigma_{12n} &= \Lambda_n^{-1}\frac{|\mathcal{K}_{1n}|}{|l_n|}\left[\sum_{i=1}^n\sum_{j\neq i}\int\int w_n(\boldsymbol{x})w_n(\boldsymbol{y})f_i(\boldsymbol{x})f_j(\boldsymbol{y})\sigma_\Psi(\boldsymbol{x}-\boldsymbol{y})\sum_{\boldsymbol{l}\in l_n}\mathbb{1}\{\boldsymbol{x},\boldsymbol{y}\in B_n(\boldsymbol{l};\boldsymbol{0})\}\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\right]\Lambda_n^{-1}\\
&= \Lambda_n^{-1}\frac{|\mathcal{K}_{1n}|}{|l_n|}\left[\sum_{i=1}^n\sum_{j\neq i}\int_{\|\boldsymbol{h}\|\le b_n}\sigma_\Psi(\boldsymbol{h})\int_{R_{2n}}w_n(\boldsymbol{y}+\boldsymbol{h})w_n(\boldsymbol{y})f_i(\boldsymbol{y}+\boldsymbol{h})f_j(\boldsymbol{y})\right.\\
&\quad\left.\times\sum_{\boldsymbol{l}\in l_n}\mathbb{1}\{\boldsymbol{y}+\boldsymbol{h},\boldsymbol{y}\in B_n(\boldsymbol{l};\boldsymbol{0})\}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{h}\right]\Lambda_n^{-1}(1+o(1))\\
&= \int\sigma_\Psi(\boldsymbol{h})Q(\boldsymbol{h})\mathrm{d}\boldsymbol{h}(1+o(1)).
\end{aligned}\tag{3.56}$$

By (3.55) and (3.56), we have shown (3.53), which completes the proof.

□

Proof of Theorem 3.5. The proof of this theorem extends the one discussed by Lahiri and Zhu (2006) to the proposed sampling design. For convenience, denote $\Phi(\cdot;\Sigma)$ to be the probability measure of $N(\mathbf{0},\Sigma)$. Based on (C9) and the Taylor's expansion, we have

$$\begin{aligned} 0 &= \sum_{k\in\mathcal{K}_n}\{S_n^*(\boldsymbol{k};\boldsymbol{t})-\hat{c}_n(\boldsymbol{k})\} \\ &= \sum_{k\in\mathcal{K}_n}\{S_n^*(\boldsymbol{k};\hat{\boldsymbol{\beta}}_n)-\hat{c}_n(\boldsymbol{k})\}+\Lambda_n\Gamma_n\lambda_n(\boldsymbol{t}-\hat{\boldsymbol{\beta}}_n)\chi_0+R_n^*(\boldsymbol{t}), \end{aligned} \tag{3.57}$$

where $\boldsymbol{t}$ is a solution of the first equality, $\Gamma_n=\sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}$, and $R_n^*(\boldsymbol{t})$ is obtained by subtraction. To be more specific, we have $R_n^*(\boldsymbol{t})=\{R_{1n}^*(\boldsymbol{t})+R_{2n}^*(\boldsymbol{t})+R_{3n}^*(\boldsymbol{t})\}\Lambda_n(\boldsymbol{t}-\hat{\boldsymbol{\beta}}_n)$, where

$$\begin{aligned} R_{1n}^*(\boldsymbol{t}) &= \sum_{i=1}^n \boldsymbol{w}(\boldsymbol{s}_i)\boldsymbol{w}(\boldsymbol{s}_i)^{\mathrm{T}}\Lambda_n^{-1}-\sum_{k\in\mathcal{K}_n}\sum_{i=1}^n \boldsymbol{w}(\boldsymbol{s}_i)\boldsymbol{w}(\boldsymbol{s}_i)^{\mathrm{T}}\Lambda_n^{-1}\mathbb{1}\{\boldsymbol{s}_i\in B_n(I_{\boldsymbol{k}};\boldsymbol{k})\}, \\ R_{2n}^*(\boldsymbol{t}) &= \sum_{k\in\mathcal{K}_n}\sum_{i=1}^n \boldsymbol{w}(\boldsymbol{s}_i)\boldsymbol{w}(\boldsymbol{s}_i)^{\mathrm{T}}\Lambda_n^{-1}\mathbb{1}\{\boldsymbol{s}_i\in B_n(I_{\boldsymbol{k}};\boldsymbol{k})\}\int_0^1[\Psi'\{\hat{Z}(\boldsymbol{s}_i)-u\boldsymbol{w}(\boldsymbol{s}_i)^{\mathrm{T}}(\boldsymbol{t}-\hat{\boldsymbol{\beta}}_n)\}-\Psi'\{\hat{Z}(\boldsymbol{s}_i)\}]\mathrm{d}u, \\ R_{3n}^*(\boldsymbol{t}) &= \sum_{k\in\mathcal{K}_n}\sum_{i=1}^n \boldsymbol{w}(\boldsymbol{s}_i)\boldsymbol{w}(\boldsymbol{s}_i)^{\mathrm{T}}\Lambda_n^{-1}\mathbb{1}\{\boldsymbol{s}_i\in B_n(I_{\boldsymbol{k}};\boldsymbol{k})\}\left[\Psi'\{\hat{Z}(\boldsymbol{s}_i)\}-E\Psi'\{Z(\mathbf{0})\}\right] \\ &= \sum_{k\in\mathcal{K}_n}\sum_{i=1}^n \boldsymbol{w}(\boldsymbol{s}_i)\boldsymbol{w}(\boldsymbol{s}_i)^{\mathrm{T}}\Lambda_n^{-1}\mathbb{1}\{\boldsymbol{s}_i\in B_n(I_{\boldsymbol{k}};\boldsymbol{k})\}\left[\Psi'\{Z(\boldsymbol{s}_i)\}-E\Psi'\{Z(\mathbf{0})\}\right]+o_p(1), \end{aligned}$$

where the second equality of $R_{3n}^*(\boldsymbol{t})$ holds by (C9), (3.38), (3.39) and Theorem 3.3. Besides, based on (3.38) and (3.39), we have

$$\Gamma_n = H+o_p(1), \qquad \text{a.s. } (P_{\boldsymbol{S}}). \tag{3.58}$$

By a similar argument in the proof of Theorem 2 (Lahiri and Zhu, 2006) and lemma 3.9, we have, for any $\epsilon_0>0$,

$$P_{\cdot|\boldsymbol{S}}\left(\sup_{B\in\mathcal{C}}\left|P_*\left[\Lambda_n^{-1}\sum_{k\in\mathcal{K}_n}\{S_n^*(\boldsymbol{k};\hat{\boldsymbol{\beta}}_n)-\hat{c}_n(\boldsymbol{k})\}\in B\right]-\Phi(B;\Sigma_\beta)\right|>\epsilon_0\right)=o(1), \qquad \text{a.s. } (P_{\boldsymbol{S}}). \tag{3.59}$$

Now, it remains to prove, for any $\epsilon_n\downarrow 0$,

$$P_{\cdot|\boldsymbol{S}}\left(P_*\left[\left\|\Lambda_n^{-1}\{R_{1n}^*(\boldsymbol{t})+R_{2n}^*(\boldsymbol{t})+R_{3n}^*(\boldsymbol{t})\}\right\|>\epsilon_n\right]>\epsilon_0\right)=o(1), \qquad \text{a.s. } (P_{\boldsymbol{S}}). \tag{3.60}$$

First, we show

$$E_{\cdot|\boldsymbol{S}}E_*\left\|\Lambda_n^{-1}\{R_{1n}^*(\boldsymbol{t})+R_{2n}^*(\boldsymbol{t})+R_{3n}^*(\boldsymbol{t})\}\right\|=o\,(1) \tag{3.61}$$

with some $\boldsymbol{t}$ such that $\left\|\Lambda_n(\boldsymbol{t}-\hat{\boldsymbol{\beta}}_n)\right\| = O(1)$.

First, consider $\Lambda_n^{-1}R_{1n}^*(\boldsymbol{t})$.

$$\begin{aligned}
& E_*\left[\sum_{\boldsymbol{k}\in\mathcal{K}_{1n}}\sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\mathbb{1}(\boldsymbol{s}_i\in R_{2n})\mathbb{1}\{\boldsymbol{s}_i\in B(I_{\boldsymbol{k}};\boldsymbol{k})\}\right] \\
= \; & E_*\left[\sum_{i=1}^n\sum_{\boldsymbol{k}\in\mathcal{K}_{1n}} \boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\mathbb{1}(\boldsymbol{s}_i\in R_{2n})\mathbb{1}\{\boldsymbol{s}_i\in B(I_{\boldsymbol{k}};\boldsymbol{k})\}\right] \\
= \; & \sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\mathbb{1}(\boldsymbol{s}_i\in R_{2n})\left|\mathcal{K}_{1n}\right|\frac{b_n^d(1+o(1))}{|l_n|} \\
= \; & \sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\mathbb{1}(\boldsymbol{s}_i\in R_{2n})(1+o(1)). \qquad (3.62)
\end{aligned}$$

Besides, by (C1) and (C7), we have

$$\sum_{\boldsymbol{k}\in\mathcal{K}_{2n}}\sum_{i=1}^n \|\boldsymbol{v}_i\|\,\mathbb{1}\{\boldsymbol{s}_i\in B(I_{\boldsymbol{k}};\boldsymbol{k})\} + \sum_{\boldsymbol{k}\in\mathcal{K}_{1n}}\sum_{i=1}^n \|\boldsymbol{v}_i\|\,\mathbb{1}(\boldsymbol{s}_i\notin R_{2n})\mathbb{1}\{\boldsymbol{s}_i\in B(I_{\boldsymbol{k}};\boldsymbol{k})\} = o(1). \qquad (3.63)$$

Thus, by (3.38), (3.39), (3.62) and (3.63), we have

$$E_*\left[\sum_{\boldsymbol{k}\in\mathcal{K}_n}\sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\mathbb{1}\{\boldsymbol{s}_i\in B(I_{\boldsymbol{k}};\boldsymbol{k})\}\right] = \Gamma_n + o(1). \qquad (3.64)$$

Denote $\boldsymbol{e}_i$ to be the vector such that all the elements are 0 except that its i-th one is 1, and $i = 1,\ldots,n$. For any $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$,

$$\begin{aligned}
V_*\left[\boldsymbol{e}_i^{\mathrm{T}}\sum_{\boldsymbol{k}\in\mathcal{K}_n}\sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\mathbb{1}(\boldsymbol{s}_i\in B(I_{\boldsymbol{k}};\boldsymbol{k}))\boldsymbol{e}_j\right] \; & = \; |\mathcal{K}_n|\,V_*\left[\boldsymbol{e}_i^{\mathrm{T}}\sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\mathbb{1}(\boldsymbol{s}_i\in B(I_{\boldsymbol{k}};\boldsymbol{k}))\boldsymbol{e}_j\right] \\
& \leq \; C\,|\mathcal{K}_n|\,E_*\left[\sum_{i=1}^n \|\boldsymbol{v}_i\|^4\,\mathbb{1}(\boldsymbol{s}_i\in B(I_{\boldsymbol{k}};\boldsymbol{k}))\right] \\
& \leq \; C\,|\mathcal{K}_n|\sum_{i=1}^n \|\boldsymbol{v}_i\|^4\,b_n^d/\,|l_n| \\
& = \; o(1), \qquad (3.65)
\end{aligned}$$

where $\boldsymbol{e}_i$ is a vector of 0 but 1 for the i-th element, C is a constant, and the last equality holds by (C3) and (C7).

Thus, by (3.64) and (3.65), we have

$$E_*\left\|\Gamma_n^{-1}R_{1n}^*(\boldsymbol{t})\right\| = o(1). \qquad (3.66)$$

Next, we consider $\Lambda_n^{-1} R_{2n}^*(\boldsymbol{t})$. Since $\|\Lambda_n^{-1} R_{2n}^*(\boldsymbol{t})\| \leq \sum_{\boldsymbol{k}\in\mathcal{K}_n} \sum_{i=1}^n \|\boldsymbol{v}_i\|^{2+\gamma} \mathbb{1}\{\boldsymbol{s}_i \in B_n(I_{\boldsymbol{k}}; \boldsymbol{k})\} \left\|\Lambda_n(\boldsymbol{t}-\hat{\boldsymbol{\beta}}_n)\right\|^\gamma$, we have

$$E_* \left\|\Lambda_n^{-1} R_{2n}^*(\boldsymbol{t})\right\| = o\,(1)\,. \tag{3.67}$$

where the result holds based on (3.37), and recall that $\left\|\Lambda_n(\boldsymbol{t}-\hat{\boldsymbol{\beta}}_n)\right\| = O(1)$.

Now, we consider $\Gamma_n^{-1} R_{3n}^*(\boldsymbol{t})$. For simplicity, denote $W_{jl}(\boldsymbol{s}_i) = \boldsymbol{e}_j^{\mathrm{T}} \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{e}_l [\Psi'\{Z(\boldsymbol{s}_i)\} - \chi_0]$ for $j, l = 1, \ldots, p$.

$$\begin{aligned}
& E_{\cdot|\boldsymbol{S}} \left(V_* \left[\sum_{\boldsymbol{k}\in\mathcal{K}_n} \sum_{i=1}^n W_{jl}(\boldsymbol{s}_i) \mathbb{1}\{\boldsymbol{s}_i \in B(I_{\boldsymbol{k}}; \boldsymbol{k})\} \right] \right) \\
\leq\;& E_{\cdot|\boldsymbol{S}} \left(E_* \left[\sum_{\boldsymbol{k}\in\mathcal{K}_n} \sum_{i=1}^n W_{jl}(\boldsymbol{s}_i) \mathbb{1}\{\boldsymbol{s}_i \in B(I_{\boldsymbol{k}}; \boldsymbol{k})\} \right]^2 \right) \\
=\;& |l_n|^{-1}\, E_{\cdot|\boldsymbol{S}} \left(\sum_{\boldsymbol{k}\in\mathcal{K}_n} \sum_{\boldsymbol{x}\in l_n} \left[\sum_{i=1}^n W_{jl}(\boldsymbol{s}_i) \mathbb{1}\{\boldsymbol{s}_i \in B(\boldsymbol{x}; \boldsymbol{k})\} \right]^2 \right) \\
=\;& o(1),
\end{aligned} \tag{3.68}$$

where the last equality holds based on the result in Lemma 2 of Lahiri and Zhu (2006) by setting $m_n = b_n^d$ based on (C3).

Thus, by (3.66), (3.67) and (3.68), we have (3.61) holds. Therefore, we have

$$\left\|\Lambda_n^{-1} R_n^*(\boldsymbol{t})\right\| \leq o(1) \left\|\Lambda(\boldsymbol{t}-\hat{\boldsymbol{\beta}}_n)\right\| \tag{3.69}$$

for some $\boldsymbol{t}$ such that $\left\|\Lambda(\boldsymbol{t}-\hat{\boldsymbol{\beta}}_n)\right\| = O(1)$.

By Markov's inequality, we can prove (3.60). Together with (3.59), Theorem 3.5 is proved. □

References

Antal, E. and Tillé, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population, *Journal of the American Statistical Association* **106**(494): 534–543.

Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*, Springer Science & Business Media, New York.

Babu, G. J. and Singh, K. (1984). On one term edgeworth correction by efron's bootstrap, *Sankhyā́: The Indian Journal of Statistics, Series A (1961-2002)* **46**(2): 219–232.

Babu, G. J. and Singh, K. (1985). Edgeworth expansions for sampling without replacement from finite populations, *Journal of Multivariate Analysis* **17**(3): 261–278.

Beaumont, J. F. and Patak, Z. (2012). On the generalized bootstrap for sample surveys with special attention to poisson sampling, *International Statistical Review* **80**(1): 127–148.

Bickel, P. J. and Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling, *The Annals of Statistics* **12**(2): 470–482.

Booth, J. G., Butler, R. W. and Hall, P. (1994). Bootstrap methods for finite populations, *Journal of the American Statistical Association* **89**(428): 1282–1289.

Chao, M.-T. and Lo, S.-H. (1985). A bootstrap method for finite population, *Sankhyā: The Indian Journal of Statistics, Series A* **47**(3): 399–405.

DiCiccio, T. J. and Romano, J. P. (1995). On bootstrap procedures for second-order accurate confidence limits in parametric models, *Statistica Sinica* **5**(1): 141–160.

Dong, Q., Elliott, M. R. and Raghunathan, T. E. (2014). A nonparametric method to generate synthetic populations to adjust for complex sampling design features, *Survey Methodology* **40**(1): 29–46.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *The Annals of Statistics* **7**(1): 1–26.

Feller, W. (2008). *An introduction to probability theory and its applications*, Vol. 2, John Wiley, New York.

Fuller, W. A. (2009). *Sampling Statistics*, John Wiley, Hoboken.

Gross, S. (1980). Median estimation in sample surveys, *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 181–184.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer Science & Business Media, New York.

Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations, *The Annals of Mathematical Statistics* **14**(4): 333–362.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association* **47**(260): 663–685.

Klenke, A. (2014). *Probability Theory: A Comprehensive Course*, Universitext, second edn, Springer-Verlag London Ltd., London.

Mashreghi, Z., Haziza, D. and Léger, C. (2016). A survey of bootstrap methods in finite population sampling, *Statistics Surveys* **10**: 1–52.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data, *Journal of the American Statistical Association* **83**(401): 231–241.

Shao, J. and Sitter, R. R. (1996). Bootstrap for imputed survey data, *Journal of the American Statistical Association* **91**(435): 1278–1288.

Sitter, R. R. (1992a). A resampling procedure for complex survey data, *Journal of the American Statistical Association* **87**(419): 755–765.

Sitter, R. R. (1992b). Comparing three bootstrap methods for survey data, *Canadian Journal of Statistics* **20**(2): 135–154.

Sverchkov, M. and Pfeffermann, D. (2004). Prediction of finite population totals based on the sample distribution, *Survey Methodology* **30**(1): 79–92.

CHAPTER 4. SPATIO-TEMPORAL BALANCED SAMPLING DESIGN FOR LONGITUDINAL AREA SURVEYS

Zhonglei Wang and Zhengyuan Zhu

Abstract

A spatially balanced sample, which spreads over the study region well, can produce optimal estimates of annual finite population quantities when the study variable is weakly dependent over the spatial region with respect to a super-population model. In this paper, we propose a spatio-temporal balanced sampling design such that the sample for each year is spatially balanced, and that combined from consecutive years is also spatially balanced. We propose design-based estimators of the annual status and change, and the corresponding variance estimators are also derived. Simulation studies show that the spatial balance of a sample generated by the proposed spatio-temporal balanced sampling design is good, and design-based estimators work well.

The proposed spatio-temporal balanced sampling design is applied to the National Resources Inventory rangeland on-site survey based on the study variable "average soil aggregate stability" of Texas from 2009 to 2013, and better estimators of the annual status can be achieved compared with the generalized estimation equation model, which is currently used by the rangeland on-site survey, based on the original sample. Although the spatio-temporal balanced sampling design is proposed in a two-dimensional space, it can be generalized to higher dimensions easily.

Key Words: Annual change; Annual status; Environment survey; Regression estimator; Variance estimator.

4.1 Introduction

Annual study about the rangeland conditions, such as the rangeland health and soil surface aggregate stability, is one of the main tasks of the rangeland on-site survey, which is a spatio-temporal survey conducted by the Natural Resources Conservation Service of the U.S. Department of Agriculture as a part of the National Resources Inventory. The rangeland on-site survey focuses on some important issues, such as the changes in rangeland conditions, and such information is essential for policy makers and scientists. Bellhouse (1977) showed that a spatially balanced sample can produce optimal estimates under a spatial setting in a two-dimensional space. However, due to the budget limit, the current sampling design of the National Resources Inventory is a stratified two-stage design (Nusser and Goebel, 1997; Breidt and Fuller, 1999), by which a sample with good spatial balance can hardly be generated for the rangeland study.

Spatially balanced sampling designs have been investigated during the past thirty years. Stevens and Olsen (2004) proposed a generalized random tessellation stratified design to generate a sample via a systematic sampling process after partitioning and mapping the two-dimensional locations into a line by a hierarchical permutation and a random quadrant-recursive map, and they showed that a sample generated by the proposed sampling design is more spatially balanced than that generated by independent random sampling or spatially stratified sampling. Theobald et al. (2007) proposed a reversed randomized quadrant-recursive raster method and showed that the sample generated by their proposed design is more spatially balanced than the one generated by the generalized random tessellation stratified design. However, it is difficult to extend the sampling designs discussed by Stevens and Olsen (2004) and Theobald et al. (2007) to a higher dimensional space. There are other sampling designs by partitioning the study region; see Munholland and Borkowski (1996), Breidt (1995), and Dunn and Harrison (1993) for details. Lister and Scott (2009) proposed to generate a sample by mapping two-dimensional locations into a line using the Peano curve, and they demonstrated that the proposed sampling design performs similarly as the generalized random tessellation stratified design in terms of spatial balance. Instead of reducing the dimensionality, Grafström et al. (2012) generalized the pivotal method (Deville and Tille, 1998) to get a sample by updating the inclusion probabilities of two nearby locations in the finite population until all inclusion probabilities are updated to 0 or 1, and they showed that a sample generated by the proposed local pivotal method is more spatially balanced than that by the generalized random tessellation stratified design. Another advantage of the local pivotal method is that it can be easily generalized to a higher dimensional space. Based on the local pivotal method, Grafström et al. (2017) proposed a double sampling procedure to generate a spatially balanced sample by incorporating the distribution of auxiliary variables. Benedetti et al. (2015) discussed and compared some spatially balanced sampling designs.

Cochran (1977) and Scott (1998) showed that a repeated survey, which revisits the sample in the first round, can improve the precision of an annual change estimate, while a non-repeated survey, that is, a temporarily independent survey, is more efficient in estimating the annual status. Therefore, a temporal sample with a repeated panel can produce reasonable estimates of both annual status and change. Jessen (1942) and Patterson (1950) proposed a sampling design with part replacement, which can be regarded as a panel survey; see Duncan and Kalton (1987), Nusser et al. (1998), Urquhart et al. (1998), Breidt and Fuller (1999), McDonald (2003), Schreuder et al. (2004), Fancy et al. (2009), and Lackey and Stein (2015) for details about the panel survey. Wikle and Royle (1999) proposed to use a measurement error model and a hidden Markov process to study the best dynamic design for environmental studies under a certain criterion.

In this paper, we propose a spatio-temporal balanced sampling design with a repeated panel such that the sample from each year is spatially balanced, and that combined from consecutive years is also spatially balanced. A repeated panel is included so that information can be borrowed when estimating annual quantities. Based on the proposed spatio-temporal balanced sampling design, we also investigate design-based regression estimators of the annual status and change, and the corresponding variance estimators are derived as well.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the rangeland on-site survey. We propose a spatio-temporal balanced sampling design with a repeated panel in Section 3. The design-based estimators of the annual status and change are investigated in Section

4. Simulation studies are conducted to test the performance of the proposed sampling design and design-based estimators in Section 5. Section 6 compares the proposed method with the current design and estimators used for the rangeland on-site survey. Conclusions and discussions are given in Section 7.

4.2 Rangeland on-site survey

Rangeland is defined as a land cover/use category on which the climax or potential plant cover consists principally of many native types of lands providing forage for various domestic livestocks as well as wild animals (United States Department of Agriculture, 2014). Besides, the rangeland is also valued by its wide environmental functions to many essential ecosystems, including clean water and recreation services. Since 2004, the Natural Resources Conservation Service has conducted the rangeland on-site survey based on the 17 western states, including those from North Dakota to Texas and west, and some limited data has also been collected in Louisiana and Florida. Information about some key issues in rangeland science, including the rangeland health, non-native plant species, non-native invasive plant species, bare ground, inter-canopy gaps, and soil surface aggregate stability, are collected and analyzed.

The current sampling design of the rangeland on-site survey is based on a stratified two-phase design to obtain the 1997 National Resources Inventory sample. The strata of the 1997 sample consist of political (sub-township or parish) or geographical (polygons defined by geographic coordinates) units to guarantee spatial balance, to help the data collection procedure, and to provide a mechanism for determining the sample sizes relative to the survey objectives and the heterogeneity of the natural resources. The primary sampling unit is a land tract of about 160 acres, which is defined by political or geographical boundaries, and the secondary sampling units are points; see Nusser and Goebel (1997) and Nusser et al. (1998) for details. Typically, three points are selected from each primary sampling unit according to a restricted randomization procedure to achieve good spatial balance (Nusser and Goebel, 1997). The 1997 sample consists of about 300,000 primary sampling units and about 800,000 sample points across the U.S.. There are many features of the

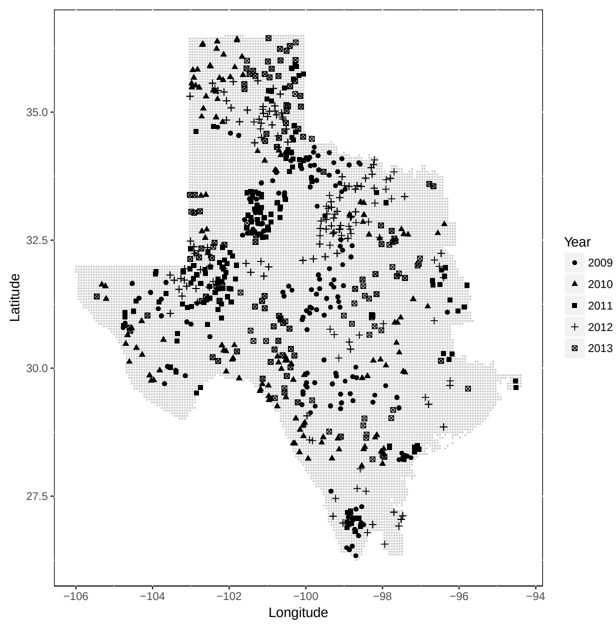

Figure 4.1: The current sampled locations of the rangeland on-site survey in Texas from 2009 to 2013. Different shapes represent the annual sample for different years, and the gray area contains 15,906 grid points in the counties that have been sampled or those in interior parts of Texas.

National Resources Inventory (U.S. Department of Agriculture, 2015). Using points allows the National Resources Inventory to provides many important issues of the rangeland science, such as the soil type and properties, soil erosion, irrigation, and cropping history. Since the National Resources Inventory is conducted based on a scientific and rigorous survey design, it is important to guarantee the integrity and confidentiality of the data gathering sites.

Figure 4.1 shows the annual samples of the rangeland on-site survey in Texas from 2009 to 2013, and the annual sample size is around 300. Although the stratified two-stage sampling design used to get the 1997 sample guarantees that the generated sample from each phase is spatially balanced, the spatial balance of the annual sample itself is not good. For example, the 2012 sample is clustered, and there are no samples in the west-south part of Texas. Thus, it is desirable to improve the current sampling design to achieve better spatial balance for the rangeland on-site survey.

4.3 Spatio-Temporal Balanced Sampling Design

Denote U to be a finite population consisting of N locations in a study region. We are interested in generating annual samples $\{A_t : t = 1, \ldots, T\}$ from U for consecutive T years with a fixed annual sample size n, such that the sample for each year is spatially balanced, and that combined from consecutive years is also spatially balanced, where A_t is the annual sample for the t-th year. Denote $\pi(\boldsymbol{s}) \in (0, 1)$ to be a pre-defined value for $\boldsymbol{s} \in U$, and it is free of the year index t. First, we demonstrate the following conditions.

(C1) The sample size satisfies $Tn = o(N)$.

(C2) The pre-defined values $\{\pi(\boldsymbol{s}) : \boldsymbol{s} \in U\}$ satisfy $\pi(\boldsymbol{s}) = O(nN^{-1})$ for $\boldsymbol{s} \in U$, and $\sum_{\boldsymbol{s}\in U} \pi(\boldsymbol{s}) = n$.

Condition (C1) guarantees that the sample size is negligible compared with the population size, and a similar condition is used by Theobald et al. (2007). Condition (C2) regulates inclusion probabilities to be homogeneous across the study region. For example, if the inclusion probabilities are the same across the study region, that is, $\pi(\boldsymbol{s}) = nN^{-1}$ for $\boldsymbol{s} \in U$, then both conditions hold if the population size is large enough. Conditions (C1)–(C2) are used to approximate the inclusion probabilities of the proposed spatio-temporal balanced sampling design by $\{\pi(\boldsymbol{s}) : \boldsymbol{s} \in U\}$. Specifically, by (C1)–(C2), we would show

$$\text{pr}(\boldsymbol{s} \in A_t) \approx \pi(\boldsymbol{s}) \quad (\boldsymbol{s} \in U, t = 1, \ldots, T), \tag{4.1}$$

where $\text{pr}(\cdot)$ is the notation for probability, and $\{A_t : t = 1, \ldots, T\}$ are the annual samples obtained by the proposed spatio-temporal balanced sampling design.

The local pivotal method is one of the most effective methods to generate a spatially balanced sample by updating the inclusion probabilities of two nearby locations successively (Grafström et al., 2012). If the (updated) inclusion probability of a location equals to 1 or 0, then that location is called to be finished and is selected into the sample or not. A finished location is removed and would not be regarded as a neighbor to other locations in the updated population. The update procedure is continued until all locations are finished. The following shows a basic procedure of a local pivotal method using $\{\pi(\boldsymbol{s}) : \boldsymbol{s} \in U\}$ as the inclusion probabilities satisfying (C2).

Step 1. Update the population by removing the finished locations.

Step 2. Randomly choose a location $\boldsymbol{s}_1$ from the updated population.

Step 3. From the updated population, choose a location $\boldsymbol{s}_2$, which is a nearest neighbor of $\boldsymbol{s}_1$.

Step 4. If $\boldsymbol{s}_1$ is not a nearest neighbor of $\boldsymbol{s}_2$ in the updated population, then go back to Step 2. Otherwise, update the inclusion probabilities of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ by the following.

- If $\pi(\boldsymbol{s}_1) + \pi(\boldsymbol{s}_2) < 1$,

$$(\pi(\boldsymbol{s}_1)', \pi(\boldsymbol{s}_2)') = \begin{cases} (0, \pi(\boldsymbol{s}_1) + \pi(\boldsymbol{s}_2)) & \text{with probability } \pi(\boldsymbol{s}_2)/\{\pi(\boldsymbol{s}_1) + \pi(\boldsymbol{s}_2)\} \\ (\pi(\boldsymbol{s}_1) + \pi(\boldsymbol{s}_2), 0) & \text{with probability } \pi(\boldsymbol{s}_1)/\{\pi(\boldsymbol{s}_1) + \pi(\boldsymbol{s}_2)\} \end{cases}.$$

- If $\pi(\boldsymbol{s}_1) + \pi(\boldsymbol{s}_2) \geq 1$,

$$(\pi(\boldsymbol{s}_1)', \pi(\boldsymbol{s}_2)') = \begin{cases} (1, \pi(\boldsymbol{s}_1) + \pi(\boldsymbol{s}_2) - 1) & \text{with probability } \{1 - \pi(\boldsymbol{s}_2)\}/\{2 - \pi(\boldsymbol{s}_1) - \pi(\boldsymbol{s}_2)\} \\ (\pi(\boldsymbol{s}_1) + \pi(\boldsymbol{s}_2) - 1, 1) & \text{with probability } \{1 - \pi(\boldsymbol{s}_1)\}/\{2 - \pi(\boldsymbol{s}_1) - \pi(\boldsymbol{s}_2)\} \end{cases}.$$

Step 5. Go back to Step 1 unless all locations are finished.

Step 1 is not listed as a single step by Grafström et al. (2012), but we use it to highlight that the finished locations are removed to obtain an updated population. If there are more than one nearest neighbors for $\boldsymbol{s}_1$, we pick $\boldsymbol{s}_2$ randomly among them. Note that if $\boldsymbol{s}_2$ is a nearest neighbor of $\boldsymbol{s}_1$, it does not necessarily imply that $\boldsymbol{s}_1$ is also a nearest neighbor of $\boldsymbol{s}_2$. Although the restriction on the mutual neighborhood of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ improves the spatial balance of the sample, it is computationally intensive, and the expected computation number to obtain a sample is $O(N^3)$ in a worst case. Grafström et al. (2012) also proposed a simpler local pivotal method, and the two inclusions probabilities are always updated in Step 4 without the restriction that the two locations should be mutually nearest neighbors, and the expected computation number is $O(N^2)$ at best. See Grafström et al. (2012) for details about the local pivotal methods.

Based on the local pivotal method above, we propose the following hierarchical local pivotal method using $\{\pi(\boldsymbol{s}) : \boldsymbol{s} \in U\}$ satisfying (C2) as inclusion probabilities, and the simpler local pivotal method should be used when the population size N is large.

Step 1. Obtain a sample A of size LTn by a local pivotal method using $\pi^{(0)}(\boldsymbol{s}) = LT\pi(\boldsymbol{s})$ as the inclusion probability for $\boldsymbol{s} \in U$, where $L \geq 1$ is a pre-defined value.

Step 2. From A, use a local pivotal method to get a sample A_t for $t = 1, \ldots, T$ sequentially. Specifically, obtain A_t based on the following updated inclusion probabilities, that is,

$$\pi^{(t)}(\boldsymbol{s}) = \frac{n}{LTn - (t-1)n} \delta_{A \setminus (\cup_{0 \leq i \leq t-1} A_i)}(\boldsymbol{s}), \tag{4.2}$$

where $A_0 = \emptyset$, $A \setminus B = \{x \in A : x \notin B\}$ for two sets A and B, and $\delta_A(x) = 1$ if $x \in A$ and 0 otherwise.

In Step 1, the pre-defined value L is used to achieve good spatial balance of the sample generated by the proposed hierarchical local pivotal method. Although it is computationally efficient to set $L = 1$, the spatial balance of the annual sample is compromised since the "left-over" points are limited for choosing the annual samples for the last several years. For example, the inclusion probability for choosing A_T is either 0 or 1 conditional on A and $\{A_t : t = 1, \ldots, T-1\}$ when $L = 1$. On the other hand, if L is large, then the average distance between two nearest neighbors in A decreases, and the combined sample may not have a good spatial balance since two adjacent points may be selected into two different annual samples in Step 2. In Section 5, we conduct a simulation study to demonstrate the effect of L on the spatial balance of a sample generated by the proposed hierarchical local pivotal method. Based on (C1)–(C2), we have $T^{-1} \min_{\boldsymbol{s} \in U} \pi(\boldsymbol{s})^{-1} > 1$ when the population size is large, and we recommend that the population size should be large and $L = \min\{2, T^{-1} \min_{\boldsymbol{s} \in U} \pi(\boldsymbol{s})^{-1}\}$ in practice. In Step 2, we treat A as another finite population and use the modified inclusion probabilities in (4.2) to get the annual sample sequentially. Specifically, A_1 is obtained from A by a local pivotal method with $\{\pi^{(1)}(\boldsymbol{s}) : \boldsymbol{s} \in A\}$ being inclusion probabilities, and A_t is obtained from A with $\{\pi^{(t)}(\boldsymbol{s}) : \boldsymbol{s} \in A\}$ being inclusion probabilities for $2 \leq t \leq T$. Notice that $\pi^{(t)}(\boldsymbol{s}) = 0$ if $\boldsymbol{s} \in \cup_{1 \leq i \leq t-1} A_i$.

Based on the proposed hierarchical local pivotal method, it can be shown that $\mathrm{pr}(\boldsymbol{s} \in A_t) = \pi(\boldsymbol{s})$ for $t = 1, \ldots, T$. That is,

$$
\begin{aligned}
\mathrm{pr}(\boldsymbol{s} \in A_1) &= \mathrm{pr}(\boldsymbol{s} \in A)\mathrm{pr}(\boldsymbol{s} \in A_1 \mid \boldsymbol{s} \in A) = LT\pi(\boldsymbol{s})\frac{n}{LTn} = \pi(\boldsymbol{s}), \\
\mathrm{pr}(\boldsymbol{s} \in A_t) &= \mathrm{pr}(\boldsymbol{s} \in A)\left[\prod_{i=1}^{t-1} \mathrm{pr}(\boldsymbol{s} \notin A_i \mid \boldsymbol{s} \in A, \boldsymbol{s} \notin \cup_{j \le i-1} A_j)\right] \\
&\quad \times \mathrm{pr}(\boldsymbol{s} \in A_t \mid \boldsymbol{s} \in A, \boldsymbol{s} \notin \cup_{1 \le j \le t-1} A_j) \\
&= LT\pi(\boldsymbol{s})\left[s\prod_{i=1}^{t-1} \frac{LTn - in}{LTn - (i-1)n}\right] \frac{n}{LTn - (t-1)n} = \pi(\boldsymbol{s}) \quad (2 \le t \le T).
\end{aligned}
$$

Thus, the proposed hierarchical local pivotal method preserves the inclusion probabilities.

To borrow information for estimating the annual quantities, the t-th year sample A_t should consist of two panels, including a repeated panel (C) and a non-repeated panel (N_t) specific for the t-th year. We propose the following spatio-temporal balanced sampling design with a repeated panel to make the sample in C and N_t spatially balanced. That is,

Step 1. Sample a repeated panel. Denote $\pi_C(\boldsymbol{s}) = p\pi(\boldsymbol{s})$ to be the inclusion probability of $\boldsymbol{s} \in U$ for choosing C and use a local pivotal method to obtain a sample C of size np based on $\{\pi_C(\boldsymbol{s}) : \boldsymbol{s} \in U\}$, where $p \in (0, 1)$ is a pre-defined proportion of the repeated panel such that np is a positive integer.

Step 2. Sample non-repeated panels. Denote $\pi_N(\boldsymbol{s}) = (1-p)n\pi(\boldsymbol{s})\delta_{U_N}(\boldsymbol{s})[\sum_{\boldsymbol{s} \in U_N} \pi(\boldsymbol{s})]^{-1}$ to the the inclusion probability for $\boldsymbol{s} \in U_N$ for choosing the non-repeated panels, where $U_N = U \setminus C$. Use the proposed hierarchical local pivotal method to generate the non-repeated panels $\{N_t : t = 1, \ldots, T\}$ based on $\{\pi_N(\boldsymbol{s}) : \boldsymbol{s} \in U_N\}$.

If $p \in (0, 1)$, we have

$$
\pi_N(\boldsymbol{s}) = [1 - p\pi(\boldsymbol{s})]\frac{(1-p)n\pi(\boldsymbol{s})}{\sum_{\boldsymbol{s} \in U_N} \pi(\boldsymbol{s})} \approx (1-p)\pi(\boldsymbol{s}) \quad (\boldsymbol{s} \in U_N), \tag{4.3}
$$

where the approximation holds by (C1)–(C2). Notice that $\pi_N(\boldsymbol{s}) = (1-p)\pi(\boldsymbol{s})$ if $\pi(\boldsymbol{s}) = N^{-1}n$ for $\boldsymbol{s} \in U$. By Step 1 of the proposed spatio-temporal balanced sampling design and (4.3), we have shown (4.1).

By (C1), the generated non-repeated panels $\{N_t : t = 1, \dots, T\}$ can be approximately viewed as if they are generated from U based on the inclusion probability $(1-p)\pi(\boldsymbol{s})$ for $\boldsymbol{s} \in U$, so the repeated panel C is approximately independent of the non-repeated panels $\{N_t : t = 1, \dots, T\}$. For two different locations $\{\boldsymbol{s}_1, \boldsymbol{s}_2\} \subset U$ and two different years $\{l, k\} \subset \{1, \dots, T\}$,

$$\begin{aligned} \mathrm{cov}[\delta_{N_l}(\boldsymbol{s}_1), \delta_{N_k}(\boldsymbol{s}_2)] &= \mathrm{pr}(\boldsymbol{s}_1 \in N_l, \boldsymbol{s}_2 \in N_k) - \mathrm{pr}(\boldsymbol{s}_1 \in N_l)\mathrm{pr}(\boldsymbol{s}_2 \in N_k) \\ &\approx (1-p)\pi(\boldsymbol{s}_1)[\mathrm{pr}(\boldsymbol{s}_2 \in N_k \mid \boldsymbol{s}_1 \in N_l) - (1-p)\pi(\boldsymbol{s}_2)], \end{aligned} \tag{4.4}$$

where the approximation of the second line holds by (4.3). When Tn is fixed and L is large, $\mathrm{pr}(\boldsymbol{s}_2 \in N_k \mid \boldsymbol{s}_1 \in N_l) \approx \mathrm{pr}(\boldsymbol{s}_2 \in N_k) \approx (1-p)\pi(\boldsymbol{s}_2)$. Thus, a large value of L weakens the dependence among the non-repeated panels.

4.4 Design-Based Estimates of the Annual Status and Change

Denote $\{z_{t,\boldsymbol{s}} : t = 1, \dots, T; \boldsymbol{s} \in U\}$ to be a realization of the study variable of interest with respect to a super-population model, and we are interested in estimating the annual status and change, that is,

$$\mu_t = \frac{1}{N} \sum_{\boldsymbol{s} \in U} z_{t,\boldsymbol{s}} \quad (t = 1, \dots, T), \quad \Delta\mu_t = \mu_{t+1} - \mu_t \quad (t = 1, \dots, T-1).$$

Consider the case $p \in (0, 1)$. Denote the observations of year t in the repeated panel C to be $\{x_{t,\boldsymbol{s}} : \boldsymbol{s} \in C\}$ and those in N_t to be $\{y_{t,\boldsymbol{s}} : \boldsymbol{s} \in N_t\}$. First, we estimate μ_t by borrowing information from annual samples of other years. Let $\boldsymbol{x}_{t,\boldsymbol{s}} = (1, x_{t,\boldsymbol{s}})^{\mathrm{T}}$ for $\boldsymbol{s} \in C$ and $\boldsymbol{y}_{t,\boldsymbol{s}} = (1, y_{t,\boldsymbol{s}})^{\mathrm{T}}$ for $\boldsymbol{s} \in N_t$, where $\boldsymbol{A}^{\mathrm{T}}$ is the transpose of a matrix $\boldsymbol{A}$. Consider the following regression estimator (Fuller, 2009§2.2), that is,

$$\hat{\mu}_{t,l} = \bar{x}_t^{(HT)} + \left(\bar{\boldsymbol{\mu}}_l^{(b)} - \bar{\boldsymbol{x}}_l^{(HT)}\right)^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{t,l}, \tag{4.5}$$

where $l \in \{1, 2, \dots, T\} \setminus \{t\}$, $\bar{x}_t^{(HT)} = \sum_{\boldsymbol{s} \in C} x_{t,\boldsymbol{s}}[Np\pi(\boldsymbol{s})]^{-1}$ is the Horvitz–Thompson estimator (Horvitz and Thompson, 1952) of μ_t by the sample in the repeated panel, $\bar{\boldsymbol{x}}_l^{(HT)} = \sum_{\boldsymbol{s} \in C} \boldsymbol{x}_{l,\boldsymbol{s}}[Np\pi(\boldsymbol{s})]^{-1}$, $\hat{\boldsymbol{\beta}}_{t,l}$ is an estimate of the regression parameter with the following form

$$\hat{\boldsymbol{\beta}}_{t,l} = \left(\sum_{\boldsymbol{s} \in C} \frac{\boldsymbol{x}_{l,\boldsymbol{s}} \boldsymbol{x}_{l,\boldsymbol{s}}^{\mathrm{T}}}{p\pi(s)}\right)^{-1} \sum_{\boldsymbol{s} \in C} \frac{\boldsymbol{x}_{l,\boldsymbol{s}} x_{t,\boldsymbol{s}}}{p\pi(s)}.$$

Let $\bar{\boldsymbol{\mu}}_l^{(b)}$ be an estimator of $(1, \mu_l)^{\mathrm{T}}$ based on the observations $\{x_{l,\boldsymbol{s}} : \boldsymbol{s} \in C\} \cup \{y_{l,\boldsymbol{s}} : \boldsymbol{s} \in N_l\}$, and it is obtained by

$$\bar{\boldsymbol{\mu}}_l^{(b)} = \boldsymbol{W}_l \bar{\boldsymbol{x}}_l^{(HT)} + (\boldsymbol{I} - \boldsymbol{W}_l)\bar{\boldsymbol{y}}_l^{(HT)}, \tag{4.6}$$

where $\bar{\boldsymbol{y}}_l^{(HT)} = \sum_{\boldsymbol{s}\in N_l} \boldsymbol{y}_{l,\boldsymbol{s}}[N(1-p)\pi(\boldsymbol{s})]^{-1}$, $\boldsymbol{I}$ is the identity matrix,

$$\boldsymbol{W}_l = \hat{\boldsymbol{V}}\left(\bar{\boldsymbol{x}}_l^{(HT)}\right)^{-1}\left[\hat{\boldsymbol{V}}\left(\bar{\boldsymbol{x}}_l^{(HT)}\right)^{-1} + \hat{\boldsymbol{V}}\left(\bar{\boldsymbol{y}}_l^{(HT)}\right)^{-1}\right]^{-1},$$

and $\hat{\boldsymbol{V}}(\cdot)$ is a variance estimator with respect to the local pivotal method (Grafström and Schelin, 2014); see Section 4.8.1 for details. We use $(1-p)\pi(\boldsymbol{s})$ to approximate the inclusion probability of the sample in the non-repeated panel, but the estimation bias of $\bar{\boldsymbol{\mu}}_l^{(b)}$ is negligible under (C1)–(C2).

Denote $\boldsymbol{u}_t = (\hat{\mu}_{t,1}, \ldots, \hat{\mu}_{t,T})^{\mathrm{T}}$, where $\hat{\mu}_{t,t} = \sum_{\boldsymbol{s}\in N_t} y_{t,\boldsymbol{s}}[N(1-p)\pi(\boldsymbol{s})]^{-1}$. Since the repeated panel C is approximately independent with the non-repeated panel N_t for $t = 1, \ldots, T$, the covariance matrix for $\boldsymbol{u}_t$, say $\hat{\boldsymbol{V}}_{\boldsymbol{u},t}$, can be estimated by a function of $\hat{\boldsymbol{B}}_t = (\hat{\boldsymbol{\beta}}_{t,k_1}, \ldots, \hat{\boldsymbol{\beta}}_{t,k_{J-1}})$. Specifically,

$$\hat{\boldsymbol{V}}_{\boldsymbol{u},t} = \boldsymbol{V}_{\boldsymbol{u},t}(\hat{\boldsymbol{B}}_t) = (\hat{v}_{\boldsymbol{u},i,j}^{(t)}), \tag{4.7}$$

where $\hat{v}_{\boldsymbol{u},i,j}^{(t)}$ is the element in the i-th row and j-th column of $\hat{\boldsymbol{V}}_{\boldsymbol{u},t}$, $\hat{v}_{\boldsymbol{u},t,t}^{(t)} = \mathrm{var}(\hat{\mu}_{t,t})$, $\hat{v}_{\boldsymbol{u},l,t}^{(t)} = \hat{v}_{\boldsymbol{u},t,l}^{(t)} = 0$ for $l \in \{1, \ldots, T\} \setminus \{t\}$, and

$$\begin{aligned} \hat{v}_{\boldsymbol{u},k,l}^{(t)} &= \mathrm{cov}\left[\bar{x}_t^{(HT)} + \left(\boldsymbol{x}_k^{(HT)}\right)^{\mathrm{T}}(\boldsymbol{W}_k - \boldsymbol{I})\hat{\boldsymbol{\beta}}_{t,k}, \bar{x}_t^{(HT)} + \left(\boldsymbol{x}_l^{(HT)}\right)^{\mathrm{T}}(\boldsymbol{W}_l - \boldsymbol{I})\hat{\boldsymbol{\beta}}_{t,l}\right] \\ &\quad + \delta_{\{k\}}(l)\mathrm{var}\left[\left(\boldsymbol{y}_k^{(HT)}\right)^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{W}_k)\hat{\boldsymbol{\beta}}_{t,k}\right] \quad (k, l \in \{1, \ldots, T\} \setminus \{t\}), \end{aligned}$$

$\delta_{\{k\}}(l) = 1$ if $l = i$ and 0 otherwise, and $\mathrm{cov}(X)$ and $\mathrm{var}(X)$ are the covariance and variance of a random variable X with respect to the local pivotal method. Since the locations for the repeated panel are fixed, we can use the standard formula $\mathrm{cov}(X+Y) = [\mathrm{var}(X+Y) - \mathrm{var}(X) - \mathrm{var}(Y)]/2$ to obtain the covariance; see Section 4.8.2 for details.

Since every component in $\boldsymbol{u}_t$ is (approximately) unbiased for μ_t, we can estimate μ_t by

$$\hat{\mu}_t = \hat{\boldsymbol{w}}_{\boldsymbol{u},t}^{\mathrm{T}} \boldsymbol{u}_t, \tag{4.8}$$

where $\hat{\boldsymbol{w}}_{\boldsymbol{u},t}$ is a weight vector that minimizes the estimated variance, that is,

$$\hat{\boldsymbol{w}}_{\boldsymbol{u},t} = \arg\min_{\boldsymbol{w}} \boldsymbol{w}^{\mathrm{T}} \hat{\boldsymbol{V}}_{\boldsymbol{u},t} \boldsymbol{w}, \text{ subject to } \boldsymbol{w}^{\mathrm{T}} \mathbf{1}_T = 1,$$

where $\mathbf{1}_T = (1,1,\ldots,1)^{\mathrm{T}}$ of length T. By the Lagrange multiplier method, it can be shown that $\hat{\boldsymbol{w}}_{\boldsymbol{u},t} = \hat{\boldsymbol{V}}_{\boldsymbol{u},t}^{-1}\mathbf{1}_T(\mathbf{1}_T^{\mathrm{T}}\hat{\boldsymbol{V}}_{\boldsymbol{u},t}^{-1}\mathbf{1}_T)^{-1}$, and the corresponding variance of $\hat{\mu}_t$ is estimated by

$$\hat{V}(\hat{\mu}_t) \quad = \quad (\mathbf{1}_T^{\mathrm{T}}\hat{\boldsymbol{V}}_{\boldsymbol{u},t}^{-1}\mathbf{1}_T)^{-1}. \tag{4.9}$$

Remark 2. *As pointed out by one reviewer, when the target variable equals to 0 at many locations in the study region, the proposed estimator (4.6) may perform poorly. In such a case, instead of using a random weighted factor, we can use a fixed weight or that obtained by the design effect to reduce the bias of the estimator in this case.*

We can use a similar procedure to estimate the annual change, $\Delta\mu_t$ for $t = 1,\ldots,T-1$. Consider

$$\Delta\boldsymbol{u}_t = (\hat{\mu}_{t+1,1} - \hat{\mu}_{t,1},\ldots,\hat{\mu}_{t+1,t-1} - \hat{\mu}_{t,t-1},\hat{\mu}_{t+1,t+1} - \hat{\mu}_{t,t},\hat{\mu}_{t+1,t} - \hat{\mu}_{t,t+1},\hat{\mu}_{t+1,t+2} - \hat{\mu}_{t,t+2},\ldots,\hat{\mu}_{t+1,T} - \hat{\mu}_{t,T}).$$

Notice that the notation for $\Delta\boldsymbol{u}_t$ implicitly assume that $t \in \{2,\ldots,T-2\}$, and $\Delta\boldsymbol{u}_1$ and $\Delta\boldsymbol{u}_{T-1}$ are defined similarly. By a similar argument made in Section 4.8.2, the estimated covariance matrix of $\Delta\boldsymbol{u}_j$ is

$$\hat{\boldsymbol{V}}_{\Delta\boldsymbol{u},t} = (\hat{v}^{(t)}_{\Delta\boldsymbol{u},i,j})$$

where

$$\begin{aligned}
\hat{v}^{(t)}_{\Delta\boldsymbol{u},t,t} &= \mathrm{var}\,(\hat{\mu}_{t+1,t+1}) + \mathrm{var}\,(\hat{\mu}_{t,t}) \\
\hat{v}^{(t)}_{\Delta\boldsymbol{u},t,l} &= \hat{v}^{(t)}_{\Delta\boldsymbol{u},l,t} = 0 \\
\hat{v}^{(t)}_{\Delta\boldsymbol{u},t+1,t} &= \hat{v}^{(t)}_{\Delta\boldsymbol{u},t,t+1} = -\mathrm{cov}\left[\bar{y}^{(HT)}_t, \left(\bar{\boldsymbol{y}}^{(HT)}_t\right)^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{W}_t)\hat{\boldsymbol{\beta}}_{t+1,t}\right] - \mathrm{cov}\left[\bar{y}^{(HT)}_{t+1}, \left(\bar{\boldsymbol{y}}^{(HT)}_{t+1}\right)^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{W}_{t+1})\hat{\boldsymbol{\beta}}_{t,t+1}\right] \\
\hat{v}^{(t)}_{\Delta\boldsymbol{u},t+1,t+1} &= \mathrm{var}\left[\bar{x}^{(HT)}_{t+1} - \bar{x}^{(HT)}_t + \left(\bar{\boldsymbol{x}}^{(HT)}_t\right)^{\mathrm{T}}(\boldsymbol{W}_t-\boldsymbol{I})\hat{\boldsymbol{\beta}}_{t+1,t} - \left(\bar{\boldsymbol{x}}^{(HT)}_{t+1}\right)^{\mathrm{T}}(\boldsymbol{W}_{t+1}-\boldsymbol{I})\hat{\boldsymbol{\beta}}_{t,t+1}\right] \\
&\quad +\mathrm{var}\left[\left(\bar{\boldsymbol{y}}^{(HT)}_t\right)^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{W}_t)\hat{\boldsymbol{\beta}}_{t+1,t}\right] + \mathrm{var}\left[\left(\bar{\boldsymbol{y}}^{(HT)}_{t+1}\right)^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{W}_{t+1})\hat{\boldsymbol{\beta}}_{t,t+1}\right], \\
\hat{v}^{(t)}_{\Delta\boldsymbol{u},t+1,l} &= \hat{v}^{(t)}_{\Delta\boldsymbol{u},l,t+1} = \mathrm{cov}\left[\bar{x}^{(HT)}_{t+1} - \bar{x}^{(HT)}_t + \left(\bar{\boldsymbol{x}}^{(HT)}_t\right)^{\mathrm{T}}(\boldsymbol{W}_t-\boldsymbol{I})\hat{\boldsymbol{\beta}}_{t+1,t} - \left(\bar{\boldsymbol{x}}^{(HT)}_{t+1}\right)^{\mathrm{T}}(\boldsymbol{W}_{t+1}-\boldsymbol{I})\hat{\boldsymbol{\beta}}_{t,t+1},\right. \\
&\quad \left.\bar{x}^{(HT)}_{t+1} - \bar{x}^{(HT)}_t + \left(\bar{\boldsymbol{x}}^{(HT)}_l\right)^{\mathrm{T}}(\boldsymbol{W}_l-\boldsymbol{I})(\hat{\boldsymbol{\beta}}_{t+1,l}-\hat{\boldsymbol{\beta}}_{t,l})\right], \\
\hat{v}^{(t)}_{\Delta\boldsymbol{u},k,l} &= \hat{v}^{(t)}_{\Delta\boldsymbol{u},l,k} = \mathrm{cov}\left[\bar{x}^{(HT)}_{t+1} - \bar{x}^{(HT)}_t + \left(\bar{\boldsymbol{x}}^{(HT)}_k\right)^{\mathrm{T}}(\boldsymbol{W}_k-\boldsymbol{I})(\hat{\boldsymbol{\beta}}_{t+1,k}-\hat{\boldsymbol{\beta}}_{t,k}),\right. \\
&\quad \left.\bar{x}^{(HT)}_{t+1} - \bar{x}^{(HT)}_t + \left(\bar{\boldsymbol{x}}^{(HT)}_l\right)^{\mathrm{T}}(\boldsymbol{W}_l-\boldsymbol{I})(\hat{\boldsymbol{\beta}}_{t+1,l}-\hat{\boldsymbol{\beta}}_{t,l})\right] \\
&\quad +\delta_{\{k\}}(l)\mathrm{var}\left[\left(\bar{\boldsymbol{y}}^{(HT)}_k\right)^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{W}_k)(\hat{\boldsymbol{\beta}}_{t+1,k}-\hat{\boldsymbol{\beta}}_{t,k})\right],
\end{aligned}$$

and $k, l \in \{1, \ldots, T\} \setminus \{t, t+1\}$.

The estimator of the annual change is

$$\widehat{\Delta\mu}_t = \hat{\boldsymbol{w}}_{\Delta\boldsymbol{u},t}^{\mathrm{T}} \Delta\boldsymbol{u}_t, \tag{4.10}$$

where $\hat{\boldsymbol{w}}_{\Delta\boldsymbol{u},t} = \hat{\boldsymbol{V}}_{\Delta\boldsymbol{u},t}^{-1}\mathbf{1}_T(\mathbf{1}_T^{\mathrm{T}}\hat{\boldsymbol{V}}_{\Delta\boldsymbol{u},t}^{-1}\mathbf{1}_T)^{-1}$, and the corresponding variance is estimated by

$$\hat{V}(\widehat{\Delta\mu}_t) \quad = \quad (\mathbf{1}_T^{\mathrm{T}}\hat{\boldsymbol{V}}_{\Delta\boldsymbol{u},t}^{-1}\mathbf{1}_T)^{-1}. \tag{4.11}$$

For the case $p = 0$ or $p = 1$, no information can be borrowed from other years when estimating the annual status and change. For both case, $\hat{\mu}_t = \bar{x}_t^{(HT)} + \bar{y}_t^{(HT)}$ and $\widehat{\Delta\mu}_t = \hat{\mu}_{t+1} - \hat{\mu}_t$, where $\bar{x}_t^{(HT)} = 0$ and $\bar{y}_t^{(HT)} = 0$ for $p = 0$ and $p = 1$, respectively. The variance estimators of $\hat{\mu}_t$ are the same for $p = 0$ and $p = 1$, that is, $\hat{V}(\hat{\mu}_t) = \hat{V}(\bar{x}_j^{(HT)} + \bar{y}_t^{(HT)})$, where $\hat{V}(\bar{x}_j^{(HT)} + \bar{y}_t^{(HT)})$ is estimated with respect to the local pivotal method. When $p = 0$, the annual designs are approximately independent for different years under the assumption, so the variance of $\widehat{\Delta\mu}_t$ is estimated by $\hat{V}_{p=0}(\widehat{\Delta\mu}_t) = \hat{V}(\hat{\mu}_t) + \hat{V}(\hat{\mu}_{t+1})$. When $p = 1$, the annual design is the same for different years, and the corresponding variance of $\widehat{\Delta\mu}_t$ for $p = 1$ is estimated by $\hat{V}_{p=1}(\widehat{\Delta\mu}_t) = \hat{V}(\hat{\mu}_{t+1} - \hat{\mu}_t)$.

The variance estimators (4.9) and (4.11) can be simplified if $\pi(\boldsymbol{s}) = N^{-1}n$ for $\boldsymbol{s} \in U$; see Section 4.8.3 for details.

4.5 Simulation Study

4.5.1 Spatial Balance of the Hierarchical Local Pivotal Method

Recall that the proposed sampling design is used to generate a spatio-temporal balanced sample, such that the sample for each year is spatially balanced, and the combined sample from consecutive years is also spatially balanced. We conduct a simulation study to test the spatial balance of the proposed hierarchical local pivotal method.

The study region U consists of 2,500 equally spaced grid points in a unit square $[0, 1] \times [0, 1]$. The annual sample size is fixed to be $n = 50$, and we use the proposed hierarchical local pivotal method to generate annual samples for consecutive $T = 5$ years. The inclusion probability is

$\pi(\boldsymbol{s}) = 50/2500$ for $\boldsymbol{s} \in U$. We consider $L \in \{1, 1.5, 2, 2.5, 5, 10\}$, and recall that L is pre-defined in Step 1 of the proposed hierarchical local pivotal method in Section 3.

We use the Voronoi polygon to measure the spatial balance of a sample (Olsen et al., 2012). Denote $A = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_n\}$ to be a sample of U. The Voronoi polygon Ψ_i for $\boldsymbol{s}_i \in A$ is a collection of locations that are closer to $\boldsymbol{s}_i$ than to any other $\boldsymbol{s}_j \in A \setminus \{\boldsymbol{s}_i\}$. Let $\nu_i = \sum_{\boldsymbol{s} \in \Psi_i} \pi(\boldsymbol{s})$. Since the sample size is fixed to be n, we have $\sum_{i=1}^{n} \nu_i / n = \sum_{i=1}^{n} \sum_{\boldsymbol{s} \in \Psi_i} \pi(\boldsymbol{s})/n = n/n = 1$. If the sample is spatially balanced, we have $\nu_i \approx 1$ for $i = 1, \ldots, n$. Therefore, the variance of ν_i, denoted as

$$\zeta(A) = \mathrm{var}(\{\nu_i\}),$$

is a good measure of the spatial balance of a sample A, and a sample with good spatial balance has a small value of $\zeta(A)$. Specifically in this simulation, we consider $\zeta(A_t)$ for $t = 1, \ldots, T$, and $\zeta(\cup_{t=1}^{T} A_t)$.

We conduct 2,000 Monte Carlo simulations, and Figure 4.2 summarizes the simulation results. The left panel of Figure 4.2 shows Monte Carlo mean of $\zeta(A_t)$ for $t = 1, \ldots, T$, and the right one shows that of $\zeta(\cup_{t=1}^{T} A_t)$. When $L = 1$ and $L = 1.5$, $\zeta(A_t)$ increases as t increases, and $\zeta(A_t)$ is larger than those with $L \geq 2$ for the last three years. Even though the sampling design with $L = 1.5$ can generate a sample with better annual spatial balance than that with $L = 1$ for the first several years, the spatial balance of the fifth year sample is approximately the same for both cases. The spatial balance of the annual samples is similar for the cases when $L \geq 2$, and it is generally better than that when $L = 1$ or $L = 1.5$. On the other hand, as the value of L increases, the spatial balance of the combined sample becomes worse as shown in the right panel of Figure 4.2. To sum up, a smaller value of L undermines the spatial balance of the annual samples for the last few years because of the "left-over" issue mentioned in Section 3, and a larger value of L may generate a spatio-temporal sample with worse spatial balance for the combined sample. By the results of this simulation study, we recommend $L = \min\{2, T^{-1} \min_{\boldsymbol{s} \in U} \pi(\boldsymbol{s})^{-1}\}$ in practice.

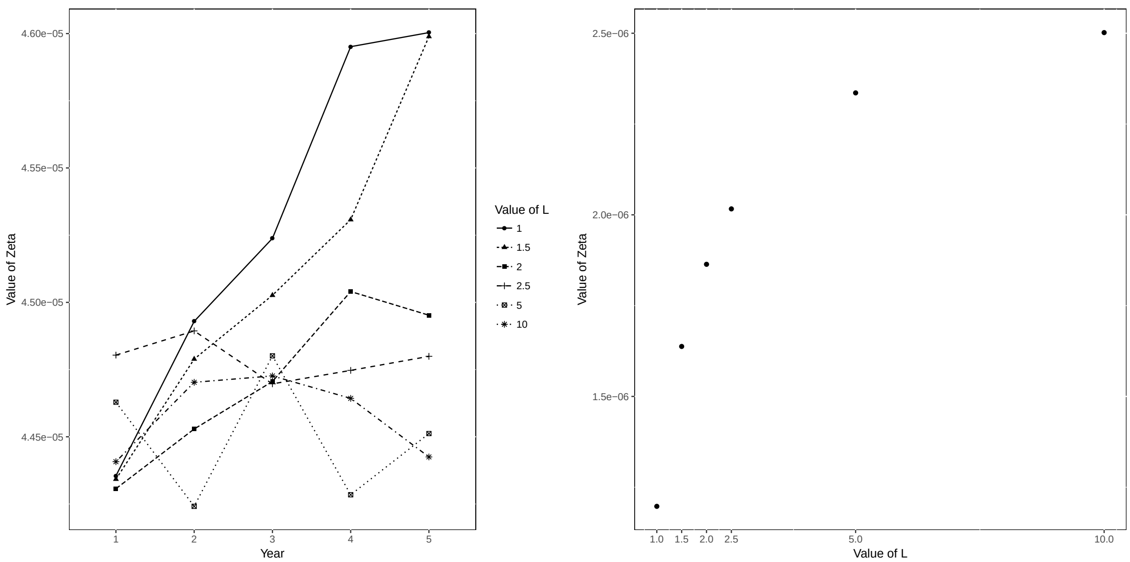

Figure 4.2: Spatial balance test of the proposed hierarchical local pivotal method. The left panel shows the spatial balance of each annual sample, that is, $\zeta(A_t)$ for $t = 1, \ldots, T$; the right panel shows the spatial balance of the combined sample, that is, $\zeta(\cup_{t=1}^{T} A_t)$.

4.5.2 Estimation Test

The performance of the proposed estimators is tested, and we set $L = 2$ for the proposed spatio-temporal balanced sampling design. The study variable at a location $\boldsymbol{s} \in U$ of the t-th year is generated by the following super-population model. That is,

$$z_{t,\boldsymbol{s}} = \nu_{t,\boldsymbol{s}} + \epsilon_{t,\boldsymbol{s}} \quad (\boldsymbol{s} \in U, \, t = 1, \ldots, T), \tag{4.12}$$

where U is the set of 100×100 equally spaced grid points in a unit square $[0, 1] \times [0, 1]$, $T = 5$, $\nu_{t,\boldsymbol{s}} = Y_t + X_{\boldsymbol{s}}$, $Y_t = \rho Y_{t-1} + \eta_t$, $\eta_t \sim N(0, 1)$ is the white noise, $X_{\boldsymbol{s}}$ is an intrinsic stationary spatial process with an exponential semi-variogram (Cressie, 2015) whose nugget, sill, and range parameters are set to be 0, 1, and ϕ, respectively, and $\epsilon_{t,\boldsymbol{s}} \sim N(0, \tau)$ is a Gaussian measurement error with τ being the standard deviation. In this simulation, we set the inclusion probability for each location to be the same, so we estimate the annual quantities by the simplified estimators shown in Section 4.8.3. We have check the relative bias of the proposed estimates shown in (4.9) and (4.11); see Section 4.8.4 for details.

We are interested in the optimal repeat panel proportion for estimating the annual status and change under different scenarios. Specifically, we consider $\rho = 0.5$ and $\rho = 0.9$ for the temporal dependence, $\phi = 0.1$ and $\phi = 0.5$ for the spatial dependence, and $\tau = 0.1$ and $\tau = 0.2$ for the measurement error. Based on one realization of the study variable generated by (4.12), we use the proposed design to obtain annual samples with $n = 150$, and the proportion of the repeated panel p ranges from 10% to 90%. For comparison reasons, we also consider two special cases, that is, $p = 0\%$ and $p = 100\%$.

For a fixed proportion p, we conduct $M = 500$ Monte Carlo simulations. Optimal proportions p are chosen such that the following objective functions are minimized under difference scenarios. That is,

$$O_1(p) = \frac{1}{5}\sum_{t=1}^{5} \hat{V}_p^{(M)}\left(\hat{\mu}_t\right), \quad O_2(p) = \frac{1}{4}\sum_{t=1}^{4} \hat{V}_p^{(M)}\left(\widehat{\Delta\mu_t}\right), \quad O_3(p) = \frac{1}{2}\left[O_1(p) + O_2(p)\right],$$

where $O_1(p)$ is associated with estimating the annual status only, $O_2(p)$ with estimating the annual change only, $O_3(p)$ is a trade-off between $O_1(p)$ and $O_2(p)$,

$$\hat{V}_p^{(M)}\left(\hat{\mu}_t\right) = \frac{1}{M}\sum_{m=1}^{M}\left(\hat{\mu}_{t,p}^{(m)}\right)^2 - \left(\frac{1}{M}\sum_{m=1}^{M}\hat{\mu}_{t,p}^{(m)}\right)^2$$

is the Monte Carlo variance for $\hat{\mu}_t$, and $\hat{V}_p^{(M)}(\widehat{\Delta\mu_t})$ is the Monte Carlo variance for $\widehat{\Delta\mu_t}$. For simplicity, we explicitly omit ϕ, ρ, and τ in the objective functions.

The optimal proportions and the corresponding minimized values of the objective functions are summarized in Table 4.1. More repeated panel is needed if the spatial dependence is high. As the measurement error increases, the optimal proportion of the repeated panel increases. This result is reasonable since the spatial trend is overwhelmed by a larger measurement error, and the information from well-spread sample becomes less. If the underlying model is additive, and we are interested in estimating the annual change, it is unnecessary to include a non-repeated panel.

The minimized values of the objective functions is compared with that under simple random sampling independently for each year in terms of the relative efficiency. That is,

$$Eff_i = \frac{V_{SRS,i}}{O_i(\hat{p}_{opt,i})} \quad (i = 1, 2, 3),$$

Table 4.1: The optimal proportion value (outside of the parentheses, unit: %) and the minimized values of the objective functions (inside of the parentheses, unit: 10^{-3}) for different scenarios. τ is the standard deviation of the measurement error in the model (4.12), and ρ and ϕ are the dependent parameters of the temporal and spatial processes, respectively. The form of the objective functions, $O_1, \ldots, O_3$, are shown in Section 5.

		$\tau = 0.1$		$\tau = 0.2$	
		$\phi = .1$	$\phi = 1$	$\phi = .1$	$\phi = 1$
O_1	$\rho = 0.1$	15 (1.021)	30 (0.406)	35 (1.527)	40 (0.712)
	$\rho = 0.9$	15 (0.995)	35 (0.413)	35 (1.527)	45 (0.712)
O_2	$\rho = 0.1$	100 (0.137)	100 (0.137)	100 (0.547)	100 (0.54)
	$\rho = 0.9$	100 (0.137)	100 (0.136)	100 (0.548)	100 (0.542)
O_3	$\rho = 0.1$	35 (0.753)	55 (0.352)	60 (1.309)	100 (0.684)
	$\rho = 0.9$	30 (0.774)	50 (0.347)	50 (1.349)	100 (0.699)

where $\hat{p}_{opt,i}$ is the optional proportion for the objective function $O_i(p)$,

$$\begin{aligned} V_{SRS,1} &= \frac{1}{5}\sum_{t=1}^{5} V_{SRS}^{(t)}, \\ V_{SRS,2} &= \frac{1}{4}\sum_{t=1}^{4} (V_{SRS}^{(t)} + V_{SRS}^{(t+1)}), \\ V_{SRS,3} &= \frac{1}{2}\left(V_{SRS,1} + V_{SRS,2}\right), \end{aligned} \tag{4.13}$$

$V_{SRS}^{(t)} = \left(1 - \frac{n}{N}\right)\frac{1}{N(N-1)}\sum_{s\in U}(z_{t,s} - \mu_t)^2$ is the variance of the estimator for μ_t under simple random sampling, and (4.13) holds since the sample is independently obtained for different years by simple random sampling. The comparison results are summarized in Table 4.2. The proposed spatio-temporal balanced sampling design is more efficient than simple random sampling under two different measurement error cases, but the gain is limited when the measurement error is large. The stronger the spatial dependence is, the more gain we get by the proposed spatio-temporal balanced sampling design.

Table 4.2: Relative efficiency of the proposed design compared with simple random sampling under different scenarios. τ is the standard deviation of the measurement error in the model (4.12), and ρ and ϕ are the dependent parameters of the temporal and spatial processes, respectively. "Eff_1" is the relative efficiency for estimating the annual status, "Eff_2" for estimating the annual change, and "Eff_3" for estimating both annual status and annual change.

		$\tau = 0.1$		$\tau = 0.2$	
		$\phi = .1$	$\phi = 1$	$\phi = .1$	$\phi = 1$
Eff_1	$\rho = 0.1$	6.58	10.72	4.53	6.39
	$\rho = 0.9$	6.75	10.54	4.53	6.39
Eff_2	$\rho = 0.1$	98.47	63.62	25.35	16.86
	$\rho = 0.9$	98.28	64.18	25.31	16.82
Eff_3	$\rho = 0.1$	13.39	18.52	7.94	9.99
	$\rho = 0.9$	13.03	18.83	7.71	9.78

4.6 Application

We use the variable "average soil aggregate stability" of Texas to propose an annual design with an optimal proportion of repeated panel for the National Resources Inventory. The real data used in this paper cannot be released in order to conserve the efficiency of the sampling design; see Section 2 for details.

We consider the real data from 2009 to 2013. In Figure 4.1, the gray area contains $N = 15,906$ grid points in the counties that have been sampled or those in interior parts of Texas, and it serves as the finite population U in this section. The size of U can be increased to get a finer grid, and the estimation results may change accordingly. The candidate proportions we consider are p =15%, 16%, ..., 85% and two specific designs with $p = 0$ and $p = 100\%$.

It is necessary to reconstruct the underlying annual trend for the average soil aggregate stability of Texas within the five years. There are different ways to construct this underlying spatio-temporal trend, such as Kriging methods for the geographic data (Cressie, 2015) and splines (Eilers and Marx, 1996; Ramsay, 2002; Lai and Wang, 2013). However, discussing the interpolation methods is not our main topic, and we use an additive model, that is, $\nu_{t,\boldsymbol{s}} = g(\boldsymbol{s}) + y_t(\boldsymbol{s})$, where $g(\boldsymbol{s})$ is estimated by the thin-plate smoothing spline (Wahba, 1990), and $y_t(\boldsymbol{s})$ is estimated by the regression spline. Thus, we can generate the finite population $z_{t,\boldsymbol{s}}$ by

$$z_{t,\boldsymbol{s}} = \nu_{t,\boldsymbol{s}} + \epsilon_{t,\boldsymbol{s}} \tag{4.14}$$

for $\boldsymbol{s} \in U$ and $t = 2009, \ldots, 2013$, where $\epsilon_{t,\boldsymbol{s}} \sim N(0, 0.3)$ is the measurement error with 0.3 being the standard deviation which is obtained from the real data. Other models can be used to generate the finite population, but the conclusion may change.

We would like to choose an optimal proportion of the repeated panel p such that the following objective function is minimized,

$$O_\alpha(p) = \alpha V_{status}(p) + (1-\alpha) V_{change}(p), \tag{4.15}$$

where $\alpha \in [0,1]$ is a trade-off between the annual status and annual change estimation, $V_{status}(p) = \frac{1}{5}\sum_{t=2009}^{2013} V_p(\hat{\mu}_t)$, $V_{change}(p) = \frac{1}{4}\sum_{t=2009}^{2012} V_p(\widehat{\Delta\mu_t})$, and the variance estimators are obtained based on (4.9) and (4.11). Notice that $\alpha = 1$ corresponds to the case where we emphasize the annual status estimation, and $\alpha = 0$ to that where we only care about the annual change estimation. We set $n = 300$ and obtain $M = 200$ replicates for each proportion p.

Figure 4.3 shows the estimation results of the objective function with different values of α. If $\alpha = 1$, the emphasis is on the annual status estimation, and the size of the repeated panel should be limited. On the other hand, if $\alpha = 0$, the emphasis is on the annual change estimation, and we should have a repeated panel with a large proportion. Based on Figure 4.3, we should choose the proportion of repeated panel to be around $p = 60\%$ when $\alpha = .5$, and this result differs if we use a different value of α.

Figure 4.4 shows one realization of 5-year annual samples for the National Resources Inventory in Texas by the proposed spatio-temporal balanced sampling design with the proportion of the repeated panel being 60%. Comparing Figure 4.4 with the original sample in Figure 4.1, we conclude that the sample generated by the proposed sampling design is more spatially balanced.

For comparison, we also use the generated finite population to compare the proposed spatio-temporal balanced sampling design and the currently used two-stage sampling design shown in

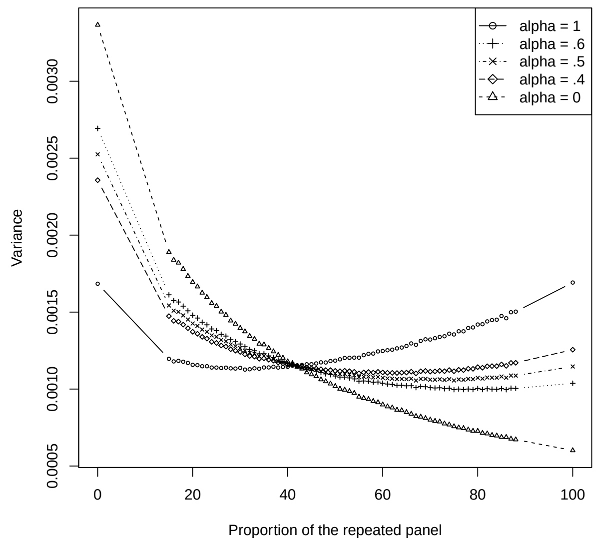

Figure 4.3: Values of the objective functions in (4.15) for different proportions of repeated panel under difference choices of α.

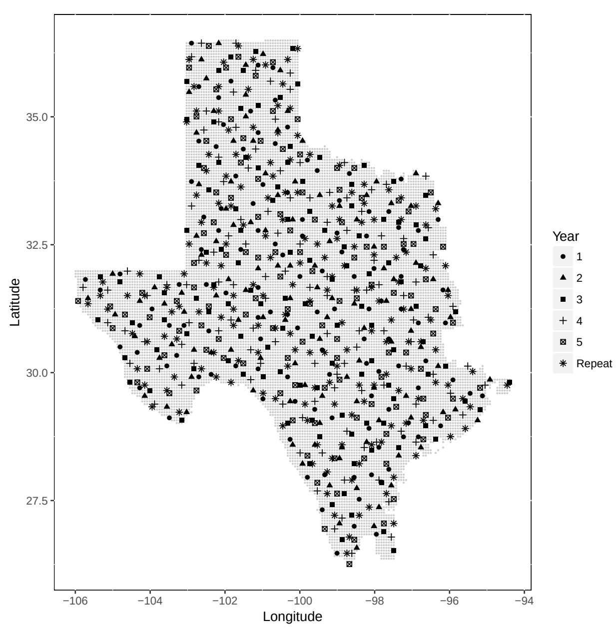

Figure 4.4: One realization of a 5-year temporal sample generated by the proposed spatio-temporal balanced sampling design for the rangeland on-site survey in Texas. The annual sample consists of the “Repeated” panel and a non-repeated panel denoted by the number of “Year”.

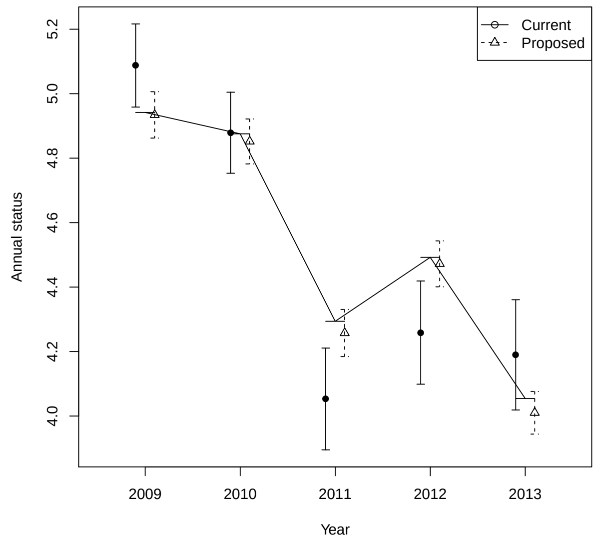

Figure 4.5: Bias of the proposed design (Proposed) and current one (Current). The horizontal line is the reference of no bias, the bias of annual status estimators are labeled by different symbols for different estimation methods, and the vertical bar shows the two standard error area around the bias.

Figure 4.1, and we set $\alpha = .5$ in (4.15) to get the optimal proportion of the repeated panel. For each location in the current sample shown in Figure 4.1, we obtain an "observation" by (4.14). The current estimation model for annual status and annual change is based on the generalized estimating equations (Hardin and Hilbe, 2002; Yan and Fine, 2004). Figure 4.5 shows the comparison result in term of the bias and the estimated standard deviation of the annual status estimators, and the proposed estimator based on the spatial-temporal spatially balanced sampling design is better than that used in the current sampling design with respect to the bias and standard deviation.

4.7 Discussions

We propose a spatio-temporal balanced sampling design with a repeated panel based on a local pivotal method. We study design-based estimates for the annual status and change, and the corresponding variance estimates are also derived. Simulation studies show that the proposed hierarchical local pivotal method can generate a sample with satisfactory spatial balance, and the optimal percentage of the repeated panel sample is discussed under different scenarios. As an application, we use the proposed spatio-temporal balanced sampling design to improve the current stratified two-stage sampling design used in the National Resources Inventory based on the information of the average soil aggregate stability in Texas, and we show that the proposed estimators are better than the current one. Although the design is proposed in a two-dimensional space, it can be easily generalized to higher dimensions.

4.8 Appendix

4.8.1 Variance Estimator of the Local Pivotal Method

We briefly review the variance estimator (Grafström and Schelin, 2014) under a local pivotal method. Without loss of generality, denote $U = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ to be a finite population. Let y be the study variable, and it takes a fixed value y_i at $\boldsymbol{s}_i$. Denote π_i to be the inclusion probability of $\boldsymbol{s}_i$ and $\sum_{i=1}^{N} \pi_i = n$.

Denote A to be a sample generated by a local pivotal method, and we observe $\{y_i : \boldsymbol{s}_i \in A\}$. Then, the mean of the study variable, that is, $\bar{y} = N^{-1} \sum_{i=1}^{N} y_i$, can be estimated by

$$\hat{y} = N^{-1} \sum_{\boldsymbol{s}_i \in A} \frac{y_i}{\pi_i}. \tag{4.16}$$

Recall that the population size is assumed to be known.

Grafström and Schelin (2014) proposed to estimate the variance of $\hat{y}$ in (4.16) by

$$\hat{V}(\hat{y}) = N^{-2} \sum_{\boldsymbol{s}_i \in A} \frac{n_i^*}{n_i^* - 1} \left(\frac{y_i}{\pi_i} - \frac{1}{n_i^*} \sum_{\boldsymbol{s}_j \in A_i^*} \frac{y_j}{\pi_j} \right)^2,$$

where A_i^* is a subset of A with n_i^* locations, $\boldsymbol{s}_i \in A_i^*$, $\boldsymbol{s}_j \in A_i^*$ if $\boldsymbol{s}_j \in A$ and $d(\boldsymbol{s}_i, \boldsymbol{s}_j) = \min_{\boldsymbol{s}_k \in A \setminus \{\boldsymbol{s}_i\}} d(\boldsymbol{s}_i, \boldsymbol{s}_k)$, and $d(\boldsymbol{s}_1, \boldsymbol{s}_2)$ is the Euclidean distance between $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

As noted by Grafström and Schelin (2014), the variance estimator of the local pivotal method works well when the underlying trend of the study variable is smooth.

4.8.2 Derivation of the Variance Estimate

For a fixed proportion $0 < p < 1$ and year $t \in \{1, 2, \ldots, T\}$, consider $l \in \{1, 2, \ldots, T\} \setminus \{t\}$, and the argument is similar for other years.

Based on the property of the spatially balanced sampling design and (C1)–(C2), we have

$$\begin{aligned} \bar{\boldsymbol{\mu}}_l^{(b)} &= (1, \mu_l)^{\mathrm{T}} + o_p(1), & (4.17) \\ \bar{\boldsymbol{x}}_l^{(HT)} &= (1, \mu_l)^{\mathrm{T}} + o_p(1). & (4.18) \end{aligned}$$

By (4.17)–(4.18), we have $\bar{\boldsymbol{\mu}}_l^{(b)} - \bar{\boldsymbol{x}}_l^{(HT)} = o_p(1)$. We also know (Fuller, 2009) that $\hat{\boldsymbol{\beta}}_{t,l} - \boldsymbol{\beta}_{t,l} = O_p(n^{-1/2})$, where $\boldsymbol{\beta}_{t,l}$ is the corresponding expectation of $\hat{\boldsymbol{\beta}}_{t,l}$, i.e.

$$\boldsymbol{\beta}_{t,l} = \left(\sum_{\boldsymbol{s} \in U} \boldsymbol{x}_{l,\boldsymbol{s}} \boldsymbol{x}_{l,\boldsymbol{s}}^{\mathrm{T}} \right)^{-1} \sum_{\boldsymbol{s} \in U} \boldsymbol{x}_{l,\boldsymbol{s}} x_{t,\boldsymbol{s}}.$$

Thus, we have

$$\bar{x}_t^{(HT)} + \left(\bar{\boldsymbol{\mu}}_l^{(b)} - \bar{\boldsymbol{x}}_l^{(HT)} \right)^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{t,l} = \bar{x}_t^{(HT)} + \left(\bar{\boldsymbol{\mu}}_l^{(b)} - \bar{\boldsymbol{x}}_l^{(HT)} \right)^{\mathrm{T}} \boldsymbol{\beta}_{t,l} + o_p(n^{-1/2}). \qquad (4.19)$$

By (4.19), the variance of the regression estimator, $\bar{x}_t^{(HT)} + \left(\bar{\boldsymbol{\mu}}_l^{(b)} - \bar{\boldsymbol{x}}_l^{(HT)} \right)^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{t,l}$, can be approximated by

$$V(\boldsymbol{\beta}_{t,l}) = \mathrm{var}\{ \bar{x}_t^{(HT)} + (\bar{\boldsymbol{\mu}}_l^{(b)} - \bar{\boldsymbol{x}}_l^{(HT)})^{\mathrm{T}} \boldsymbol{\beta}_{t,l} \}.$$

By a similar argument, we could conclude that the approximated variance estimator for $\boldsymbol{u}_t$ is a function of $\boldsymbol{B}_t = (\boldsymbol{\beta}_{t,1}, \ldots, \boldsymbol{\beta}_{t,t-1}, \boldsymbol{\beta}_{t,t+1}, \ldots, \boldsymbol{\beta}_{t,T})$. Using a plug-in estimator, we can obtain the form of (4.7).

4.8.3 The Annual Quantity Estimation based on Equal Inclusion Probability

Now we consider a special case that the inclusion probability is the same for different locations, i.e. $\pi(\boldsymbol{s}) = n/N$ for $\boldsymbol{s} \in U$. We consider to estimate μ_t by borrowing information from other years.

In this case, it can be easily shown that the weight matrix $\boldsymbol{W}_l$ in (4.6) becomes a diagonal matrix based on the variance estimator provided by Grafström and Schelin (2014). Besides, we can also derive that $\bar{\boldsymbol{x}}_l^{(HT)} = (1, \bar{x}_l)^{\mathrm{T}}$ and $\bar{\boldsymbol{y}}_l^{(HT)} = (1, \bar{y}_l)^{\mathrm{T}}$, where $(\bar{x}_l, \bar{y}_l)^{\mathrm{T}} = \left(\sum_{\boldsymbol{s} \in C} x_{l,\boldsymbol{s}}/(pn), \sum_{\boldsymbol{s} \in N_l} y_{l,\boldsymbol{s}}/(n - pn) \right)^{\mathrm{T}}$, that is, the sample mean. Thus, the regression estimator in (4.5) is simplified to be $\hat{\mu}_{t,l} = \bar{x}_t + b_{t,l}[w_l\bar{x}_l + (1-w_l)\bar{y}_l - \bar{x}_l]$, where $\bar{x}_t = \sum_{\boldsymbol{s}\in C} x_{t,\boldsymbol{s}}/(pn)$, $w_l = \mathrm{var}(\bar{x}_l)^{-1}\left[\mathrm{var}(\bar{x}_l)^{-1} + \mathrm{var}(\bar{y}_l)^{-1}\right]^{-1}$, and the regression coefficient $b_{t,l}$ has the following form

$$b_{t,l} = \frac{\sum_{\boldsymbol{s}\in C}(x_{l,\boldsymbol{s}} - \bar{x}_l)x_{t,\boldsymbol{s}}}{\sum_{\boldsymbol{s}\in C}(x_{l,\boldsymbol{s}} - \bar{x}_l)^2}.$$

Thus, we have

$$\begin{aligned}\boldsymbol{u}_t &= (\bar{x}_t + b_{t,1}[(w_1-1)\bar{x}_1 + (1-w_1)\bar{y}_1], \ldots, \bar{x}_t + b_{t,t-1}[(w_{t-1}-1)\bar{x}_{t-1} + (1-w_{t-1})\bar{y}_{t-1}], \bar{y}_t,\\ &\quad \bar{x}_t + b_{t,t+1}[(w_{t+1}-1)\bar{x}_{t+1} + (1-w_{t+1})\bar{y}_{t+1}], \ldots, \bar{x}_t + b_{t,T}[(w_T-1)\bar{x}_T + (1-w_T)\bar{y}_T])^{\mathrm{T}}.\end{aligned}$$

Based on a similar argument in Appendix 4.8.2, we can show that the variance of $\boldsymbol{u}_j$ has the following form

$$\hat{\boldsymbol{V}}_{\boldsymbol{u},t} = \boldsymbol{V}_{\boldsymbol{u},t}(\boldsymbol{b}_t) = (\hat{v}^{(t)}_{\boldsymbol{u},i,j}),$$

where $\hat{v}^{(t)}_{\boldsymbol{u},t,t} = \mathrm{var}(\hat{\mu}_{t,t})$, $\hat{v}^{(t)}_{\boldsymbol{u},l,t} = \hat{v}^{(t)}_{\boldsymbol{u},t,l} = 0$ for $l \in \{1,\ldots,T\}\setminus\{t\}$, and

$$\hat{v}^{(t)}_{\boldsymbol{u},k,l} = \mathrm{cov}[\bar{x}_t + b_{t,k}(w_k-1)\bar{x}_k, \bar{x}_t + b_{t,l}(w_l-1)\bar{x}_l] + \delta_{\{k\}}(l)b^2_{t,k}(w_k-1)^2\mathrm{var}(\bar{y}_k)$$

for $k, l \in \{1,\ldots,T\}\setminus\{t\}$.

The annual change $\Delta\mu_t$ can be estimated by

$$\begin{aligned}\Delta\boldsymbol{u}_t &= \boldsymbol{u}_{t+1} - \boldsymbol{u}_t\\ &= \begin{pmatrix} \bar{x}_{t+1} - \bar{x}_t + (w_1-1)(b_{t+1,1} - b_{t,1})(\bar{x}_1 - \bar{y}_1)\\ \vdots\\ \bar{x}_{t+1} - \bar{x}_t + (w_{t-1}-1)(b_{t+1,t-1} - b_{t,t-1})(\bar{x}_{t-1} - \bar{y}_{t-1})\\ \bar{y}_{t+1} - \bar{y}_t\\ \bar{x}_{t+1} - \bar{x}_t + b_{t+1,t}(w_t-1)(\bar{x}_t - \bar{y}_t) - b_{t,t+1}(w_{t+1}-1)(\bar{x}_{t+1} - \bar{y}_{t+1})\\ \bar{x}_{t+1} - \bar{x}_t + (w_{t+2}-1)(b_{t+1,t+2} - b_{t,t+2})(\bar{x}_{t+2} - \bar{y}_{t+2})\\ \vdots\\ \bar{x}_{t+1} - \bar{x}_t + (w_T-1)(b_{t+1,T} - b_{t,T})(\bar{x}_T - \bar{y}_T) \end{pmatrix}.\end{aligned}$$

The approximated covariance matrix of $\boldsymbol{\Delta u}_t$ is $\hat{\boldsymbol{V}}_{\Delta\boldsymbol{u},t} = (\hat{v}^{(t)}_{\Delta\boldsymbol{u},i,j})$, where $\hat{v}^{(t)}_{\Delta\boldsymbol{u},t,t} = \text{var}\,(\hat{\mu}_{t+1,t+1}) +$ $\text{var}\,(\hat{\mu}_{t,t})$, $\hat{v}^{(t)}_{\Delta\boldsymbol{u},t,l} = \hat{v}^{(t)}_{\Delta\boldsymbol{u},l,t} = 0$, $\hat{v}^{(t)}_{\Delta\boldsymbol{u},t+1,t} = \hat{v}^{(t)}_{\Delta\boldsymbol{u},t,t+1} = b_{t,t+1}(w_{t+1}-1)\text{var}(\bar{y}_{t+1}) + b_{t+1,t}(w_t-1)\text{var}(\bar{y}_t)$,

$$
\begin{aligned}
\hat{v}^{(t)}_{\Delta\boldsymbol{u},t+1,t+1} &= b^2_{t,t+1}(w_{t+1}-1)^2\text{var}(\bar{y}_{t+1}) + b^2_{t+1,t}(w_t-1)^2\text{var}(\bar{y}_t) \\
&\quad +\text{var}\left\{\bar{x}_{t+1}[1-(w_{t+1}-1)b_{t,t+1}] - \bar{x}_t[1-(w_t-1)b_{t+1,t}]\right\}, \\
\hat{v}^{(t)}_{\Delta\boldsymbol{u},t+1,l} &= \hat{v}^{(t)}_{\Delta\boldsymbol{u},l,t+1} = \text{cov}\{\bar{x}_{t+1}[1-(w_{t+1}-1)b_{t,t+1}] - \bar{x}_t[1-(w_t-1)b_{t+1,t}], \\
&\quad \bar{x}_{t+1} - \bar{x}_t + (w_l-1)(b_{t+1,l}-b_{t,l})\bar{x}_l\}, \\
\hat{v}^{(t)}_{\Delta\boldsymbol{u},k,l} &= \text{cov}[\bar{x}_{t+1} - \bar{x}_t + (w_k-1)(b_{t+1,k}-b_{t,k})\bar{x}_k, \bar{x}_{t+1} - \bar{x}_t + (w_l-1)(b_{t+1,l}-b_{t,l})\bar{x}_l] \\
&\quad +\delta_{\{k\}}(l)\text{var}[(w_k-1)(b_{t+1,k}-b_{t,k})\bar{y}_k],
\end{aligned}
$$

and $k, l \in \{1, \ldots, T\} \setminus \{t, t+1\}$.

4.8.4 Performance of the Variance Estimators

We use a similar setup of Section 4.5.2 to test the performance of the variance estimators in (4.9) and (4.11), and the dimension of the grid points is 200×200. We consider two cases for the annual sample size $n \in \{50, 150\}$, and the repeated panel proportion ranges from 20% to 80% to avoid computational issues when $n = 50$.

We conduct 500 Monte Carlo simulations and use the relative bias to assess the proposed variance estimators. That is,

$$
\text{RB} = \frac{\bar{V}_{Pro} - \hat{V}_{Sim}}{\hat{V}_{Sim}},
$$

where $\bar{V}_{Pro}$ is the Monte Carlo mean of the estimated variance for a specific annual quantity, and $\hat{V}_{Sim}$ is the Monte Carlo variance of that annual quantity. Table 4.3 summarizes the simulation results. First, the temporal dependence does not have significant influence on the variance estimator *under our simulation setup* since the results of the relative bias are approximately the same for different values of ρ. Second, as ϕ increases, the relative bias of the variance estimator for the annual change decreases for most cases. Thus, the variance estimator (4.9) preforms better when a strong spatial dependence exists. However, the proposed variance estimator (4.11) tends to underestimate the variance, and it performs better when the spatial dependence is weak. In addition, when the annual sample size is negligible compared with the population size, that is, $n = 50$ under our simulation setup, both estimators perform well.

The performance of the variance estimators (4.9) and (4.11) depends on the performance of the variance estimator of the local pivotal method, and Grafström and Schelin (2014) demonstrated that their proposed variance estimator may overestimate the variance in some cases, but its performance is satisfactory when the study variable has a smooth trend. Specifically, the variance estimator of the local pivotal method overestimate the empirical one by about 17% in their simulation study.

Table 4.3: Simulation results for the relative bias for different proportions. Based on different proportions of the repeated panel, the mean of relative biases are shown outside of the parenthesis, and their standard deviation is inside of the parenthesis.

			Annual Status					Annual Change			
			$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=1$	$t=2$	$t=3$	$t=4$
$n=50$	$\rho=0.5$	$\phi=0.1$	0.14(0.11)	0.09(0.06)	0.10(0.10)	0.10(0.12)	0.13(0.11)	-0.14(0.12)	-0.13(0.09)	-0.11(0.09)	-0.07(0.10)
		$\phi=0.5$	0.10(0.15)	0.06(0.14)	0.08(0.12)	0.06(0.10)	0.09(0.13)	-0.14(0.15)	-0.15(0.17)	-0.17(0.13)	-0.19(0.12)
		$\phi=0.1$	0.20(0.09)	0.24(0.12)	0.27(0.15)	0.23(0.13)	0.21(0.11)	-0.15(0.12)	-0.12(0.11)	-0.14(0.14)	-0.13(0.13)
		$\phi=0.5$	0.10(0.12)	0.11(0.11)	0.10(0.09)	0.09(0.14)	0.12(0.17)	-0.21(0.10)	-0.22(0.15)	-0.24(0.13)	-0.22(0.16)
	$\rho=0.9$	$\phi=0.1$	0.13(0.14)	0.14(0.14)	0.11(0.16)	0.13(0.15)	0.09(0.14)	-0.13(0.10)	-0.09(0.14)	-0.13(0.11)	-0.10(0.12)
		$\phi=0.5$	0.06(0.13)	0.09(0.09)	0.07(0.14)	0.03(0.11)	0.09(0.10)	-0.16(0.14)	-0.16(0.16)	-0.20(0.12)	-0.12(0.11)
		$\phi=0.1$	0.25(0.16)	0.25(0.13)	0.21(0.11)	0.25(0.11)	0.22(0.14)	-0.13(0.15)	-0.13(0.13)	-0.12(0.13)	-0.13(0.14)
		$\phi=0.5$	0.15(0.19)	0.07(0.13)	0.13(0.15)	0.09(0.15)	0.09(0.11)	-0.26(0.16)	-0.23(0.16)	-0.21(0.14)	-0.22(0.16)
$n=150$	$\rho=0.5$	$\phi=0.1$	0.43(0.15)	0.44(0.16)	0.40(0.17)	0.41(0.16)	0.42(0.10)	-0.04(0.10)	-0.01(0.13)	-0.01(0.10)	0.01(0.10)
		$\phi=0.5$	0.35(0.09)	0.35(0.09)	0.38(0.13)	0.38(0.15)	0.39(0.17)	-0.02(0.13)	-0.07(0.09)	-0.06(0.09)	-0.10(0.09)
		$\phi=0.1$	0.43(0.13)	0.42(0.13)	0.41(0.17)	0.38(0.14)	0.52(0.20)	-0.05(0.09)	-0.05(0.11)	-0.07(0.10)	-0.08(0.08)
		$\phi=0.5$	0.18(0.10)	0.25(0.13)	0.22(0.12)	0.19(0.16)	0.26(0.12)	-0.15(0.09)	-0.17(0.10)	-0.18(0.09)	-0.20(0.10)
	$\rho=0.9$	$\phi=0.1$	0.42(0.09)	0.43(0.09)	0.38(0.10)	0.41(0.13)	0.39(0.11)	-0.01(0.07)	-0.05(0.11)	-0.07(0.07)	-0.05(0.08)
		$\phi=0.5$	0.33(0.13)	0.33(0.09)	0.38(0.14)	0.31(0.11)	0.33(0.08)	-0.09(0.11)	-0.05(0.13)	-0.08(0.08)	-0.07(0.11)
		$\phi=0.1$	0.39(0.14)	0.36(0.11)	0.36(0.15)	0.36(0.10)	0.40(0.14)	-0.07(0.12)	-0.05(0.09)	-0.06(0.11)	-0.04(0.11)
		$\phi=0.5$	0.27(0.14)	0.28(0.12)	0.26(0.13)	0.25(0.14)	0.26(0.15)	-0.16(0.14)	-0.14(0.07)	-0.12(0.08)	-0.16(0.08)

References

Bellhouse, D. R. (1977). Some optimal designs for sampling in two dimensions, *Biometrika* **64**: 605–611.

Benedetti, R., Piersimoni, F., Postiglione, P. et al. (2015). *Sampling Spatial Units for Agricultural Surveys*, Springer, Berlin.

Breidt, F. J. (1995). Markov chain designs for one-per-stratum spatial sampling, *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, DC, pp. 356–361.

Breidt, F. J. and Fuller, W. A. (1999). Design of supplemented panel surveys with application to the National Resources Inventory, *Journal of Agricultural, Biological, and Environmental Statistics* **4**: 391–403.

Cochran, W. G. (1977). *Sampling Techniques*, John Wiley, New York.

Cressie, N. A. C. (2015). *Statistics for Spatial Data*, revised edn, John Wiley, New York.

Deville, J.-C. and Tille, Y. (1998). Unequal probability sampling without replacement through a splitting method, *Biometrika* **85**: 89–101.

Duncan, G. J. and Kalton, G. (1987). Issues of design and analysis of surveys across time, *International Statistical Review* **55**: 97–117.

Dunn, R. and Harrison, A. R. (1993). Two-dimensional systematic sampling of land use, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **42**: 585–601.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties, *Statistical Science* **11**: 89–102.

Fancy, S. G., Gross, J. E. and Carter, S. L. (2009). Monitoring the condition of natural resources in US national parks, *Environmental Monitoring and Assessment* **151**: 161–174.

Fuller, W. (2009). *Sampling Statistics*, John Wiley, Hoboken.

Grafström, A., Lundström, N. L. P. and Schelin, L. (2012). Spatially balanced sampling through the pivotal method, *Biometrics* **68**: 514–520.

Grafström, A. and Schelin, L. (2014). How to select representative samples, *Scandinavian Journal of Statistics* **41**: 277–290.

Grafström, A., Zhao, X., Nylander, M. and Petersson, H. (2017). A new sampling strategy for forest inventories applied to the temporary clusters of the swedish national forest inventory, *Canadian Journal of Forest Research* **47**(9): 1161–1167.

Hardin, J. W. and Hilbe, J. M. (2002). *Generalized estimating equations*, Chapman & Hall/CRC, London.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association* **47**: 663–685.

Jessen, R. J. (1942). Statistical investigation of a sample survey for obtaining farm facts, *Iowa Agricultural Experiment Station Research Bulletin* **304**: 54–59.

Lackey, L. G. and Stein, E. D. (2015). Evaluating alternative temporal survey designs for monitoring wetland area and detecting changes over time in California, *JAWRA Journal of the American Water Resources Association* **51**: 388–399.

Lai, M.-J. and Wang, L. (2013). Bivariate penalized splines for regression, *Statistica Sinica* **23**: 1399–1417.

Lister, A. J. and Scott, C. T. (2009). Use of space-filling curves to select sample locations in natural resource monitoring studies, *Environmental Monitoring and Assessment* **149**: 71–80.

McDonald, T. (2003). Review of environmental monitoring methods: survey designs, *Environmental Monitoring and Assessment* **85**: 277–292.

Munholland, P. L. and Borkowski, J. J. (1996). Simple latin square sampling + 1: a spatial design using quadrats, *Biometrics* **52**: 125–136.

Nusser, S., Breidt, F. and Fuller, W. (1998). Design and estimation for investigating the dynamics of natural resources, *Ecological Applications* **8**: 234–245.

Nusser, S. M. and Goebel, J. J. (1997). The National Resources Inventory: a long-term multi-resource monitoring programme, *Environmental and Ecological Statistics* **4**: 181–204.

Olsen, A. R., Kincaid, T. M. and Payton, Q. (2012). Spatially balanced survey designs for natural resources, *in* R. A. Gitzen, J. J. Millspaugh, A. B. Cooper and D. S. Licht (eds), *Design and Analysis of Long-term Ecological Monitoring Studies*, Cambridge University Press, pp. 126–150. Cambridge Books Online.

Patterson, H. D. (1950). Sampling on successive occasions with partial replacement of units, *Journal of the Royal Statistical Society. Series B (Methodology)* **12**: 241–255.

Ramsay, T. (2002). Spline smoothing over difficult regions, *Journal of the Royal Statistical Society: Series B (Methodology)* **64**: 307–319.

Schreuder, H. T., Ernst, R. L. and Ramirez-Maldonado, H. (2004). *Statistical Techniques for Sampling and Monitoring Natural Resources*, US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins.

Scott, C. T. (1998). Sampling methods for estimating change in forest resources, *Ecological Applications* **8**: 228–233.

Stevens, D. L. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources, *Journal of the American Statistical Association* **99**: 262–278.

Theobald, D. M., Stevens, D. L., White, D., Urquhart, N. S., Olsen, A. R. and Norman, J. B. (2007). Using GIS to generate spatially balanced random survey designs for natural resource applications, *Environmental Management* **40**(1): 134–146.

United States Department of Agriculture (2014). 2014 national resources inventory rangeland resource assessment, Website. Last checked: December 5, 2017.

Urquhart, N. S., Paulsen, S. G. and Larsen, D. P. (1998). Monitoring for policy-relevant regional-trends over time, *Ecological Applications* **8**: 246–257.

U.S. Department of Agriculture (2015). Summary report: 2012 national resources inventory. Natural Resources Conservation Service, Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia.

Wikle, C. K. and Royle, J. A. (1999). Space: time dynamic design of environmental monitoring networks, *Journal of Agricultural, Biological, and Environmental Statistics* pp. 489–507.

Yan, J. and Fine, J. (2004). Estimating equations for association structures, *Statistics in medicine* **23**(6): 859–874.

CHAPTER 5. CONCLUSION

In this dissertation, we use bootstrap methods to make inference for three commonly used sampling designs and a one-per-stratum sampling design. We also propose a spatio-temporal sampling design to obtain annual samples, such that the sample for each year is spatially balanced, and the one combined from consecutive years is also spatially balanced. The following summarizes the main contribution of these three parts.

Wald method is widely used to make inference in survey sampling, but it is not second-order accurate. In the first part, we propose bootstrap methods for three commonly used sampling designs, and show that the bootstrap methods are second-order accurate. Thus, the proposed bootstrap methods work better than the Wald method in making inference especially when the sample size is not large. The idea of the bootstrap methods can be extended to other sampling designs if the central limit theorem holds, but the second-order accuracy may not guarantee. In addition, the bootstrap methods can be used to do hypothesis tests in survey sampling, and this is a future research topic.

Spatially balanced sampling designs are widely used in environmental studies, but the variance estimator for the population mean estimate is hard to obtain for most such designs. Although there are some approximations, their performance is not satisfactory. In the second part, we propose to use a block bootstrap to obtain the variance estimator and make inference under a one-per-stratum sampling design. We show that the block bootstrap method is valid under weak dependent settings theoretically. Simulation study is conducted to test the performance of the block bootstrap method under different scenarios. However, the block bootstrap method is computationally intensive, and it is a future research topic to explore other block bootstrap methods to improve the computational efficiency.

Although there are many researches about the spatially balanced sampling designs, few of them are under the spatio-temporal settings. In the third part, we propose a spatio-temporal balanced sampling design to generate annual samples. In order to borrow information to estimate the annual status and change, we propose to use the regression estimators. We derive the design-based variance estimators for the estimates of the annual quantities. The proposed method is applied to the National Resources Inventory rangeland on-site survey, and it is more efficient to estimate the annual quantities.